

High Resolution Model-Independent Carbon Monoxide Emission Localization and Characterization Using TROPOMI

Master Thesis

Lodewijck R. Foorthuis

Delft University of Technology
Faculty of Aerospace Engineering
Section Aircraft Noise and Climate Effects (ANCE)



Cover page: *Steam rises from the chimneys of a coal-fired power plant in Roggendorf, Germany, on Nov. 8, 2019.* Via Federico Gambarini, Getty Images.

“I am proud to be part of a species where a subset of its members willingly put their lives at risk to push the boundaries of our existence.”

– Neil DeGrasse Tyson

High Resolution Model-Independent Carbon Monoxide Emission Localization and Characterization Using TROPOMI

Master Thesis

by

Lodewijck R. Foorthuis

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday June 10, 2022 at 10:00 AM.

Student number: 4395131
Project duration: February 8, 2021 – June 10, 2022
Thesis committee: Dr. I. C. Dedoussi, TU Delft, Thesis supervisor
Dr. Ir. J. D. Maasackers, SRON, Main supervisor
Prof. Dr. I. Aben, SRON, Overall supervisor
Prof. Dr. M. Snellen, TU Delft, Committee chair
Dr. A. Bombelli, TU Delft, Committee member

This report is confidential and cannot be made public until June 10th, 2024.

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Executive Summary

Carbon monoxide (CO) and methane (CH₄) are among the major air pollutant and greenhouse gases in the atmosphere. Accurate monitoring of the emissions of these gases plays an important role in climate change mitigation. The TROPospheric Monitoring Instrument (TROPOMI) was launched in 2017 and measures, among other, CO and methane concentrations with a high spatial resolution of $5.5 \times 7.5 \text{ km}^2$ and daily global coverage. For high accuracy CO and methane emission source localization and quantification using TROPOMI data, generally Chemical Transport Models (CTMs) are used, which are very complex and require extensive resources in terms of time and computational load. To analyse emissions without the use of a CTM, a CTM-independent method has been developed by Beirle et al. [1] for NO_x. The main goal of this research is to test the CTM-independent method by Beirle et al. [1] for CO and methane. Two separate functional models are developed for the gases, where the CO model is fully optimized and tuned.

The method uses the divergence of the flux to estimate emissions. The divergence is independent of changes in wind fields and can therefore be used for long term analyses of emissions, which makes the method suitable for the separation of closely-space emissions sources, as well as the quantification of a full grid of emissions. Using the divergence, sources are identified by an iterative Gaussian peak-fitting algorithm.

The developed divergence method for CO uses background and altitude correction, as well as a land mask filter to improve the performance of the method. Synthetic pseudo data from the WRF-chem model is used to tune the divergence method. The sensitivity analyses using the pseudo data resulted in an optimal method, which makes use of the ERA5 100 m (altitude) wind field with a filter on low ($< 1 \text{ m/s}$) and high ($> 10 \text{ m/s}$) wind speeds at a model resolution of 0.03° . The divergence model uncertainty is estimated as 20%, with a quantification limit of 50 Gg/a.

The optimised divergence model is used to test seven case studies using TROPOMI data, which showed that the model is able to separate closely-spaced emission sources up to 13 km in distance. The model successfully identified all known sources for the TROPOMI case studies. Furthermore, the model identified multiple sources that were not previously known in bottom-up inventories and previous studies. The TROPOMI case studies showed the potential of the divergence method to detect and separate emission point sources, as well as quantify CO emissions from cities. The case studies demonstrated the ability of the divergence model in detecting unknown emission sources, showing the added value of the divergence method in regions where limited information on emission source locations and quantifications is available.

Key Points

- A CTM-independent divergence method is developed and optimised for carbon monoxide which estimates emissions by divergence calculations.
- Gaussian peak-fitting is used to identify and separate closely-spaced emission sources.
- Synthetic pseudo data from the WRF-chem CTM is used to tune and optimize the model.
- Seven case studies are analysed using the optimised divergence model and TROPOMI data.

Keywords

TROPOMI, Carbon monoxide, Methane, Chemical transport model, Divergence, Peak-fitting, Source localization, Emission quantification, Source separation, Pseudo data

Preface

This report describes the thesis research project I have been working on since early February 2021. It is the final part of my masters degree in Control & Operations, Aerospace Engineering, at Delft University of Technology. The research is conducted at SRON Netherlands Institute for Space Research. Even though part of my research has been carried out remotely, everyone at SRON's Earth group made me feel like I was an important part of their research team. I am extremely satisfied with the results of my research, and I am greatly thankful for everyone at SRON who supported me during my thesis research.

I would like to thank my supervisors, Irene (TU Delft) and Bram & Ilse (SRON), without whom this project would not have been possible. I can not express the gratitude for the extensive feedback I have received on my research, literature review and thesis report. You helped me frame my research, set my goals and interpret my results, which ultimately led to this thesis report. My special thanks go out to Bram, who has been my daily supervisor during the project. Besides our weekly meetings you always took the time to answer my questions, often late at night. Your guidance has shaped the research and helped me obtain a satisfying answer to my research questions.

During my time at SRON I was not the only student doing an internship or thesis research. I want to thank all my fellow students whom I spent time with during and after working hours: Pratik, Tjomme, Maarten and Allard and all other intern students at SRON. Mostly I want to thank Berend and Pieter, who helped me with coding, writing and brainstorming. Your company and support has truly made my time at SRON special. Besides, Pieter has supplied the high-resolution PRISMA plume figure, for which I am more than grateful. Also, I want to thank Gourav, Stijn, Pankaj and Sudhanshu who supplied me with reading material, helped me with coding questions and gave me tips and tricks. My special thanks go out to Gijs, who supplied the land mask filter, pseudo data and CSF and IME quantification estimates for Cairo and Riyadh.

Lastly, I want to thank everyone who endured me talking about CO and flux calculations for the past year: my (former) roommates, study friends and the tennis crew: Jeroen, Florian, Luca and mostly Jelle, who accompanied me on long days writing this report. My special thanks go out to my family: Willem, Gerda, Maurits and Angelina, and in-laws for their genuine interest in my project and its outcome, and most of all Lisanne, who never got tired of me complaining and/or explaining anything related to my research. Your understanding and support has got me through the difficult times, and the successful finalisation of this project could not have taken place without you.

To whomever reads this report: I hope you enjoy the extensive description of my work and hope you are just as amazed by the results of this research as I am.

*Lodewijck R. Foorthuis
Delft, May 2022*

Contents

Executive Summary	III
Preface	V
List of Figures	IX
List of Tables	XVII
Nomenclature	XXI
Abbreviations	XXIII
1 Introduction	1
1.1 Introduction to the Problem	1
1.2 The Beirle Method	2
1.3 Introduction to the Report	4
2 Theoretical Background	5
2.1 Gases in the Earth's Atmosphere	5
2.1.1 Nitrogen Dioxide	6
2.1.2 Carbon Monoxide	8
2.1.3 Methane	9
2.1.4 Mixing Ratio	11
2.1.5 Lifetime	11
2.2 Sources and Sinks	12
2.2.1 Nitrogen Dioxide	12
2.2.2 Carbon Monoxide	13
2.2.3 Methane	14
2.3 Measuring Atmospheric Trace Gases	15
2.3.1 Physical Measurement Principles	15
2.4 TROPOMI	16
2.4.1 TROPOMI Trace Gas Concentration Retrievals	17
2.4.2 Data Quality and Known Data Issues	19
2.5 Emission Quantification Methods for TROPOMI	19
2.5.1 The Beirle Method	19
2.5.2 Cross-sectional Flux Method	20
2.5.3 Integrated Mass Enhancement Method	21
2.5.4 Supporting Data and Databases	21
3 Research Definition	23
3.1 Relevance of the Project	23
3.2 Research Question(s)	24
4 Methodology	25
4.1 Data Preparation	25
4.1.1 The TROPOMI Data Product	25
4.1.2 Data Selection	26

4.1.3	Pixel Oversampling	27
4.2	Model Functions	30
4.2.1	Data Corrections and Filters	30
4.3	Model Analysis	35
4.3.1	Divergence Calculations	37
4.3.2	Gaussian Peak-fitting	39
4.3.3	Emission Quantification	40
4.3.4	An Explanation For Divergence Output Form	42
4.4	Adaptation for the Methane Model	43
4.5	WRF-generated CO Pseudo Data	44
4.5.1	Sensitivity Analysis	47
5	Results	53
5.1	Optimizing the Model Parameters	53
5.1.1	TROPOMI Filtering Thresholds	54
5.1.2	Sensitivity Analyses	55
5.1.3	Multiple Source Domains	59
5.1.4	Uncertainty Estimation	60
5.1.5	Detection and Quantification Limit	60
5.1.6	Alternative Numerical Differentiation Methods	62
5.2	Case Studies	63
5.2.1	Duisburg, Germany	64
5.2.2	Hebei, China	66
5.2.3	Jharkhand, India	69
5.2.4	Cairo, Egypt	72
5.2.5	Riyadh, Saudi-Arabia	76
5.2.6	The Influence of the Covid-19 Pandemic on CO Emissions in Wuhan	82
5.2.7	Testing the Methane Model for Korpezhe, Turkmenistan	83
5.2.8	Conclusions from Case Studies	85
6	Conclusion	87
7	Recommendations	91
7.1	Recommendations for Future Research	91
7.2	Recommendations for Applications	92
	Data Availability	93
	Bibliography	95
A	Flowchart for the Divergence Method	101
B	WRF Pseudo-data Locations	103
C	Case Studies	105
C.1	Identified sources	106
C.2	Duisburg, Germany	107
C.3	Hebei, China	110
C.4	Jharkhand, India	115
C.5	Cairo, Egypt	119
C.6	Riyadh, Saudi-Arabia	123

List of Figures

1.1	Schematic illustration for the divergence of flux profiles. The figure shows emission profiles (E) for urban and industry sources. The concentration profile (C) is slightly distorted due to the wind vector (w). The divergence (D) preserves strong flux gradients around point sources [8].	2
1.2	Tropospheric NO ₂ column over Riyadh as derived from TROPOMI. (A) Single overpass on 17 December 2017. Arrows indicate wind vectors. (B) Temporal average for December 2017 to October 2018. (C) Emission map of TROPOMI for December 2017 to October 2018. Clear individual sources can be seen [1].	3
2.1	The layers of the Earth's atmosphere, classified by their vertical temperature gradient [9].	6
2.2	Daily NO ₂ concentrations as observed by UK ground stations for roadside, urban and rural areas. 1 ppb \approx 0.53 $\mu\text{g m}^{-3}$ [19].	8
2.3	Global mean carbon monoxide levels at an altitude of 3,700 meters, as measured by the MOPITT instrument [24].	9
2.4	Globally averaged, monthly mean atmospheric methane abundance determined from marine surface sites, as published by the Global Monitoring Division of NOAA's Earth System Laboratory [36].	10
2.5	Principles used in remote-sensing measurements [3].	16
2.6	The TROPOMI measurement principle, which shows the resolution of 7 km and the swath of 2600 km. All ground pixels are measured at the same time [6].	17
4.1	The FORTRAN sub-pixel oversampling method, showing four gridcells being overlapped by a TROPOMI pixel [67].	28
4.2	Two TROPOMI pixels, one with a low value (blue) and one with a high value (red) and the weighted average that is allocated to each gridcell. The white gridcell is not overlapped, and therefore results in a zero value.	29
4.3	A methane plot (left) showing missing pixels over rivers. The oversampling routine improves coverage for low resolutions (middle) and partly improves coverage for high resolutions (right).	29
4.4	Average CO concentration map (g m^{-2}) for two months in early 2019, showing India and Nepal. Extremely low CO concentration values can be seen for low surface pressure (high altitude) around the Himalayas.	31
4.5	The altitude-CO relation for India on 2019-01-01. The dots denote the CO concentration measurement at a certain altitude, the red line is the linear regression line. The <i>TROPOMI</i> CO value denotes the oversampled CO values with a fifth percentile background correction. CO concentration measurements below 1,000 m (light blue) are not used for the calculation of the altitude correction factor.	32
4.6	The seasonality of the altitude correction factor. The red line represent the fitted average.	33
4.7	A simulated case without background correction. Enhancements in the divergence map are the result of the wind speed in combination with a high background.	34

4.8	A simulation for the background correction. The erroneous divergence enhancements are mostly fixed.	34
4.9	The coverage map, showing grid coverage (average of TROPOMI pixels per gridcell) for concentration (left) and coverage for divergence (right). The data loss by individual missing pixels ('+' form) can be seen in the divergence map.	35
4.10	The daily geometric revisit frequency of TROPOMI.	36
4.11	The model output showing gridpoints with their number of pixels and the corresponding divergence. The output becomes stable above 20 pixels.	36
4.12	A cut-out from the model grid, showing the gridcells (black), the gridpoint used for the divergence calculation (red) and the gridpoints used for the numerical derivative (blue).	38
4.13	A cut-out from the model grid, showing the gridcells without a TROPOMI CO value (blank with red dot). All other gridcells do have a TROPOMI CO value. This data gap results in extensive data loss (red gridcells) for the divergence calculation. The green gridcells do have a divergence value.	38
4.14	The emission map, showing the identified sources by the Gaussian peak-fitting algorithm (left), as well as the emission estimate from the REAS inventory for Jharkhand, India. All sources match known emission sources.	41
4.15	The CO concentration map for January 31st, 2019 with the GEOS-FP 10 m wind field (left), showing multiple plumes in the Jharkhand area. The wind speed and direction is shown by the black arrows. The resulting divergence output is shown on the right. Known emission sources are marked.	42
4.16	The results of the iterative peak-fitting algorithm (left bottom) is compared to the wind rose (right bottom). Similarities can be seen in the form of both outputs. Each contour line of the Gaussian fit in the sub-figure left-below represent a drop of 0.15 with respect to the highest value of the normalized divergence data.	43
4.17	The yearly mean coverage for methane, showing high coverage in desert regions and low coverage over mountain areas.	44
4.18	To test whether enhancements in the emission output are not artefacts, the albedo (left) and aerosol (middle) map can be compared to the emission map (right). Emission sources are identified with makers, all other divergence enhancements are artefacts.	44
4.19	The averaging kernel for three different categories of cloudiness: strict cloud clearing (black), semi clear-sky (yellow) and high optical thick clouds (blue). The standard deviation is indicated as error bars [71].	46
4.20	A day in 2019 with wind speeds up to 18 m/s, resulting in strong wind-induced noise fields in the divergence map. Due to strong winds, the plumes in the grid are broken-up, resulting in dipoles in the divergence.	48
4.21	The TROPOMI across-track pixel resolution for a domain in Hebei, China, early July 2020. For each day, the large variation of TROPOMI resolution throughout the swath can be seen between 0.059° (min) and 0.41° (max).	49
4.22	A divergence analysis for a location in India on a extremely high model resolution (0.01°). The model output shows excessive stripe-like noise.	50
5.1	Average data loss (%) for specified thresholds for AOT filtering in Gent, Belgium. . . .	54
5.2	The average grid coverage for various AOT and HSL filter settings for all WRF locations, showing increased data loss for stricter filter settings.	54
5.3	Emission estimates from the divergence method compared to the actual WRF emissions [%] for specified thresholds for the AOT and HSL filtering. No clear improvement for stricter filter settings is observed.	55

5.4	The sensitivity analysis of the 9 locations (marked by colors) for the datasets, showing the best performance for PW. The AK dataset is closest to the TROPOMI data. Sources above 50 Gg/a are marked by dots, sources below 50 Gg/a are marked by a triangle.	56
5.5	Analysis results for varying wind field input, showing GEOS 10 m and ERA5 10 m and 100 m. A combination of GEOS and ERA 5 10 m winds is shown by GE5, selecting the min or max value of both wind fields. A wind speed filter (WSP) slightly improves the GEOS 10 m estimate. The best performing input is chosen as ERA5 100 m.	57
5.6	The additional sensitivity test for the wind speed filter for all individual wind fields.	57
5.7	The analysis for the TROPOMI resolution filter, showing no clear improvement for both filter settings.	58
5.8	Relative performance for varying model resolutions. The best performing resolution for emission quantification is chosen as 0.04°. For resolutions > 0.06°, not all model analyses resulted in a successful fit, which makes that these resolutions show less results.	58
5.9	The Gaussian peak-fitting sequence for different resolutions. Iteration 0 shows the initial divergence output. Iteration 1 shows the first peak-fitting procedure. Iteration 2 shows the second peak-fitting procedure. The sources can be separated up to 0.03 deg. Each contour line of the Gaussian fits in the subfigures represent a drop of 0.15 with respect to the highest value of the normalized divergence data.	60
5.10	Relative model performance for the emission estimation for all nine pseudo data locations.	61
5.11	Model performance for varying Source Strength (SS). The results converge after SS = 300 Gg/a. The results are within the bounds of uncertainty after 50 Gg/a.	61
5.12	Relative model performance for the emission estimation for all nine pseudo data locations for a second-order and fourth-order numerical method. The fourth-order outperforms the second-order method for 7 out of 9 locations.	62
5.13	The Duisburg area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°), including contours of large cities. Large CO concentration enhancements in the center of the grid can be seen.	64
5.14	Emission estimates from the divergence model per gridcell for the two identified plants in Duisburg (left), compared to estimates from the non-gridded TNO inventory for 2017 (right).	65
5.15	Quantification estimates for the Duisburg case study. The results of the divergence model are compared to different inventories. <i>* Results for the divergence method are shown for 01/July/2018 - 30/June/2021.</i>	65
5.16	The low resolution (0.5°) analysis for East-China (2020), with two main concentration enhancements over Xingtai and Hebei. The area marked by the black box is chosen for the high resolution analysis.	66
5.17	The Hebei area, showing the TROPOMI CO concentration (left) and the corresponding divergence (right) at high resolution (0.03°).	67
5.18	The Hebei area, showing the TROPOMI CO concentration for January 9th, 2019 and February 6th, 2019 at high resolution (0.03°). The black arrows indicate the ERA5 100 m wind direction and speed. Multiple plumes can be seen, which indicate multiple emission sources.	67
5.19	The Hebei area, showing the results from the peak-fitting algorithm (left). Five sources are identified (A-E). The emission estimates can be compared to the REAS inventory (right). It can be seen that neither the locations of the sources nor the emission estimates match. Both emission estimates (divergence method and REAS) show estimates per gridcell.	68

5.20	The Jharkhand area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Multiple enhancements can be spotted.	69
5.21	The Jharkhand area, showing the TROPOMI CO concentration (left) and corresponding divergence (right) at high resolution (0.03°). The markers indicate the locations of known emission sources.	70
5.22	TROPOMI CO concentration at high resolution (0.03°) for March 1st and March 21st, 2019. The black arrows indicate the ERA5 100 m wind direction and speed. Multiple plumes can be seen, which indicates multiple sources. The signal from the unknown location does show plume-like behavior, which indicates a source.	71
5.23	The identified emission sources by the Gaussian peak-fitting algorithm (left) and the REAS inventory (right). Three unmarked sources can be seen, which indicate unknown emission sources.	72
5.24	Emission quantification estimates from the divergence method, previous research by Sadavarte and Venkataraman [75] and bottom-up inventories. <i>* Results for the divergence method are shown for 01/July/2018 - 30/June/2021.</i>	72
5.25	The Cairo area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Large CO concentration enhancements in the center of the grid can be seen.	73
5.26	The CO concentration and corresponding divergence for the city of Cairo (outline). The city center shows large concentration and divergence enhancements. Multiple divergence enhancements can be seen east of the Nile river and in the south-west of the grid, possible altitude-correlated.	73
5.27	Emission estimates from the divergence method (left) and the EDGAR inventory (right). Estimates are shown as average emissions per gridcell.	74
5.28	A new model run with the possible artefact location as grid center. A non-Gaussian enhancement can be seen in the divergence.	75
5.29	Two individual days (August 10th and January 14th, 2019) showing CO enhancements, but no plume-like behaviour.	75
5.30	Emission quantification estimates for the divergence method, compared to different inventories and quantification methods. A strong underestimate is seen for EDGAR, while DACCIWA and IME/CSF are within the uncertainty of the divergence estimate. <i>* Results for the divergence method are shown for 01/July/2018 - 30/June/2021.</i>	76
5.31	The Riyadh area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Large CO concentration enhancements in the center of the grid can be seen.	77
5.32	The city of Riyadh (outline), showing the TROPOMI CO concentration (left) and the divergence (right) at high resolution (0.03°). The markers indicate emission sources identified by the NO _x model.	77
5.33	The resulting emissions estimate for the divergence method (left) and the EDGAR inventory estimate (right). The divergence output perfectly matches the city outline, while the spatial allocation of emissions for EDGAR shows a slight mismatch.	78
5.34	Emission quantification estimates for the divergence method, compared to the EDGAR emission inventory and results from the IME/CSF quantification methods. All emission estimates are within the uncertainty of the divergence estimate. <i>* Results for the divergence method are shown for 01/July/2018 - 30/June/2021.</i>	78
5.35	Mean background-corrected CO concentration (red) and estimated CO emissions (blue) for Riyadh in 2020 using a 90-day running mean. The blue band shows the uncertainty for the emission estimate. The strong dip around February might be due to coverage issues.	79

5.36	The daily mean wind speed (gray) and grid coverage (green) for the domain of the Riyadh case study in 2020. A strong seasonality for grid coverage can be seen. The wind speed varies slightly.	80
5.37	The emission (blue) and concentration (red) difference between 2020-2019 using a 90-day running mean. A positive value means larger emissions and higher concentrations in 2020. A dip in relative concentrations can be clearly seen after March 2020.	81
5.38	The Wuhan area, showing the concentration map (left) and the corresponding divergence map (right) with the contours of the city of Wuhan (centered) and smaller lakes. Three individual sources can be seen.	82
5.39	Emission quantification estimates for different sources in the Wuhan area for 2019 and 2020 (divergence method) and 2015 (REAS). The REAS estimates are divided into <i>Industry</i> and <i>Total</i> . The divergence estimates are significantly lower than the REAS estimates.	83
5.40	The Korpezhe region (Turkmenistan-West), showing methane concentration (left) and the corresponding divergence (right) for 2019 and 2020. Since no sources are quantified by the Gaussian peak-fitting algorithm, known locations of emission sources are shown by 'x' [82].	84
5.41	A PRISMA methane observation on July 21st, 2020, showing a clear plume from a ground flare at (38.56, 54.20). The red arrow denotes the GEOS-FP 10 m wind direction (personal communication with P. Bijl).	85
A.1	Flowchart for the divergence method.	101
C.1	The case study procedure, showing all necessary steps to study a location.	105
C.2	Wind rose for the Duisburg area, showing wind direction and speed.	107
C.3	The different latitudinal and longitudinal wind speeds for the Duisburg area. The '+' markers indicate outliers.	107
C.4	Average grid coverage for the Duisburg area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.	108
C.5	Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact. No artefacts are observed.	108
C.6	Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The domain has reasonable coverage year-round.	109
C.7	The Hebei area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Multiple concentration enhancements can be spotted.	110
C.8	Emission estimates (red) from the divergence method for the five plants in Hebei versus the crude iron and steel production for 2020 (blue) [73].	110
C.9	Wind rose for the Hebei area, showing wind direction and speed.	111
C.10	The different latitudinal and longitudinal wind speeds for the Hebei area. The '+' markers indicate outliers.	111
C.11	The five resulting sources from the Hebei model run. All location match a steel/power plant in the Hebei area. Images are taken from Bing maps.	112
C.12	Average grid coverage for the Hebei area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.	112

C.13 Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact. No artefacts are identified.	113
C.14 Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The monsoon season with high cloud coverage can be clearly seen during the summer months.	114
C.15 The Jharkhand area, showing the CO concentration (left) and the divergence (right) at high resolution (0.03°) without altitude correction. The markers indicate the locations of known emission sources. Multiple altitude-induced artefacts can be spotted.	115
C.16 Wind rose for the Jharkhand area, showing wind direction and speed.	115
C.17 The different latitudinal and longitudinal wind speeds for the Jharkhand area. The '+' markers indicate outliers.	116
C.18 The two unknown locations from the Jharkhand model run. (A) is identified as the Metaliks cement and power plant. (B) is still unidentified. (C) is identified as the Saluja steel plant. Images are taken from Bing maps.	116
C.19 Average grid coverage for the Jharkhand area. The coverage is a shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.	117
C.20 Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact. No artefact indicates are observed.	117
C.21 Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The monsoon season with high cloud coverage can be clearly seen during the summer months.	118
C.22 Wind rose for the Cairo area, showing wind direction and speed.	119
C.23 The different latitudinal and longitudinal wind speeds for the Cairo area. The '+' markers indicate outliers.	119
C.24 Average grid coverage for the Cairo area. The coverage is a shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.	120
C.25 Divergence and numpix values for every gridpoint. The isolated large negative divergence values around 300-350 might indicate an artefact.	120
C.26 The location of the artefact, as identified by the divergence method.	121
C.27 The numpix map for the artefact location. A clear numpix enhancement can be seen in the center of the grid.	121
C.28 Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The grid has low coverage during the winter months, with high coverage in the summer.	122
C.29 Wind rose for the Riyadh area, showing wind direction and speed.	123
C.30 The different latitudinal and longitudinal wind speeds for the Riyadh area. The '+' markers indicate outliers.	123
C.31 Average grid coverage for the Riyadh area. The coverage is a shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.	124
C.32 Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact.	124

C.33 Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The domain has good coverage year-round.	125
C.34 Mean background-corrected CO concentration (red) and estimated CO emissions (blue) for Riyadh in 2019 using a 90-day running mean. The blue band shows the uncertainty for the emission estimation.	126
C.35 The average wind speeds and grid coverage for the Riyadh 2019 analysis. A strong seasonality for grid coverage can be seen. The average wind speed only slightly differs.	126

List of Tables

2.1	Global means for selected trace gases.	11
2.2	The different lifetimes for selected trace gases.	12
2.3	Global annual NO _x emissions [47].	13
2.4	Global annual sinks for NO _x [47].	13
2.5	Global annual CO sources [48].	14
2.6	Global annual sinks for carbon monoxide [48].	14
2.7	Bottom-up global annual sources for methane [49].	15
2.8	Global annual sinks for methane [49].	15
2.9	TROPOMI data products for individual observations [6].	18
2.10	CO qa-value parameter definition.	19
2.11	CH ₄ qa-value parameter definition.	19
4.1	The extracted variables from the TROPOMI Level 2 orbit files.	25
4.2	qa-values for corresponding AOT and HSL values.	26
4.3	Two examples to the input file.	26
4.4	Parameters that need to be defined before a model run.	27
5.1	The WRF pseudo-data locations that are used for the sensitivity analyses.	53
5.2	Additional sources for the multi-source domain sensitivity analysis.	59
5.3	Results for the sensitivity tests of the two numerical methods.	62
5.4	Locations for the case studies with the main reasons for including the domains in this research.	63
5.5	Known sources in the Duisburg area. Three steel plants are identified [70].	64
5.6	Five identified sources for the Hebei area with the locations and emission estimates from the divergence method using TROPOMI CO data, including estimates for iron and steel production [73].	68
5.7	Locations and emission estimates for five known steel plants in the Jharkhand area [75].	70
5.8	Three methane emission sources as localized by the divergence method (lat, lon). Since the model was not able to successfully fit a Gaussian to the data, no emission estimates are included. The sources are identified using research from Irakulis-Loitxate et al. [82].	84
5.9	The case studies with demonstrated performance conclusions.	85
B.1	All locations included in the WRF-chem pseudo data. Only non-coastal domains are used for the analysis.	103
C.1	Identified emission sources by the divergence method for selected case studies	106

Nomenclature

Physics Constants

g	gravitational acceleration	9.80665 m s^{-2}
M_{air}	mean molecular weight of air	$28.96 \times 10^{-3} \text{ kg mol}^{-1}$
M_{CO}	molecular weight of carbon monoxide	$28.01 \times 10^{-3} \text{ kg mol}^{-1}$
p_{atmos}	sea level standard pressure	$101,325 \text{ Pa} = 1,013.25 \text{ hPa}$
R	radius of the Earth	$6.378 \times 10^6 \text{ m}$

Divergence method symbols

A	gridcell area
A_p	overlapping area between TROPOMI pixel and gridcell
B	burden
D	divergence
E	emissions
F	flux
H	height of Gaussian
h	surface elevation
L	ratio NO_2/NO_x
L_{GIL}	global integrated loss
L_p	plume length
l	local sink
M_{res}	model resolution
n	number of molecules
P_f	performance factor
p	surface pressure
Q	source rate
Q_{div}	emission estimation by the divergence method
$RMSE$	root-mean-square error

S	sinks
S_p	pixel area
S_s	sum of squared residuals
T_{res}	TROPOMI resolution
t	time
U	wind speed
U_{eff}	effective wind speed
V	TROPOMI tropospheric column
w	wind field
z	height scattering layer
α	local loss frequency
β	altitude correction factor
ϵ	mean absolute error
θ	rotation angle (clockwise)
λ	longitude
ν	coverage
τ	aerosol optical thickness
τ_c	time constant
τ_g	global lifetime
τ_p	residence time in the plume
φ	latitude
Ψ	weighted average column
$\bar{\Omega}$	weighted average of the TROPOMI value

Other symbols

CO	carbon monoxide
CO ₂	carbon dioxide
CH ₄	methane
NO _x	nitrogen oxides
μ	mean value of a dataset
σ	standard deviation of a dataset

Definitions

a priori	"from the earlier"
GWP	Global Warming Potential, "The time-integrated Radiative Forcing due to a pulse emission of a given component, relative to a pulse emission of an equal mass of CO ₂ ." [2]
XCO	atmospheric carbon monoxide total column dry mixing ratio [ppb], "The total number of CO molecules in a vertical column above a unit surface divided by the corresponding total number of molecules of dry air in that column." [3]

Abbreviations

AK	Averaging Kernel
AOT	Aerosol Optical Thickness
CP	Cement Plant
CSF	Cross-Sectional Flux
CSV	Comma-Separated Value
CTM	Chemical Transport Model
DACCIWA	Dynamics-Aerosol-Chemistry-Cloud Interactions in West Africa
DO	Double Orbit
E-PRTR	European Pollutant Release and Transfer Register
EC	European Community
EDGAR	Emissions Database for Global Atmospheric Research
ERA5	ECMWF ReAnalysis v5
ESA	European Space Agency
GB	GigaByte
GEM	Global Energy Monitor
GEOS-FP	Goddard Earth Observing System - Forward Processing
GHGs	Greenhouse Gases
GMAO	Global Modeling Assimilation Office
GMES	Global Monitoring of the Environment
GWP	Global Warming Potential
HC	HydroCarbon
HOBIT	Harmonic Oscillator hybrid FIT
HSL	Height Scattering Layer
IME	Integrated Mass Enhancement
LEO	Low Earth Orbit
MB	MegaByte
MWIR	Medium-Wave InfraRed
NaN	Not-a-Number
NIR	Near-InfraRed
OMI	Ozone Monitoring Instrument
PBL	Planetary Boundary Layer
PP	Power Plant

ppb	parts-per-billion
ppm	parts-per-million
PRISMA	Hyperspectral PRecursor of the Application Mission
PW	Pressure Weighted
REAS	Regional Emission inventory in ASia
RMSE	Root-Mean-Square Error
SCIAMACHY	Scanning Imaging Absorption spectroMeter for Atmospheric CartograpHY
Sentinel-5P	Sentinel-5 Precursor
SO	Single Orbit
SRON	Space Research Organisation Netherlands
Suomi NPP	Suomi National Polar-orbiting Partnership
SWIR	Short-Wave InfraRed
TROPOMI	TROPOspheric Monitoring Instrument
UV	Ultra-Violet
VIIRS	Visible/Infrared Imager and Radiometer Suite
WISCO	Wuhan Iron and Steel Corporation
WRF	Weather Research and Forecasting
WSP	Wind SPeed filter

1

Introduction

This report describes the thesis research of the project “High Resolution Model-Independent Carbon Monoxide Emission Localization and Characterization Using TROPOMI”. The project is supervised by Dr. I. C. Dedoussi, assistant professor at Aircraft Noise and Climate Effects (ANCE) at Delft University of Technology, faculty of Aerospace Engineering. It is conducted at SRON Netherlands Institute for Space Research under supervision of Dr. Ir. J. D. Maasackers (daily supervisor) and Prof. Dr. I. Aben (overall supervisor).

1.1. Introduction to the Problem

From the 1800s on, humankind has made large technological advancements that still greatly influence present day society. These technological advancements have led to, besides many great societal improvements, a rapid increase of greenhouse gas emissions, which cause global warming. Some important atmospheric species for air quality and climate control are carbon monoxide (CO), methane (CH₄) and nitrogen dioxide (NO₂) [2]. As the global mean temperature rises, it gets increasingly important to accurately monitor and control the emission sources of greenhouse gases and other pollutants [4]. For this reason, it is important that reliable measurements are available. One way to accurately measure atmospheric concentrations is by using space-based remote-sensing instruments. One of these instruments is the Tropospheric Monitoring Instrument (TROPOMI). TROPOMI was launched in 2017 and obtains daily atmospheric trace gas concentrations on a spatial scale of $3.5 \times 7 \text{ km}^2$ for NO₂ and $5.5 \times 7 \text{ km}^2$ for CO and methane. Furthermore, the instrument is able to effectively detect large point sources [5].

SRON Netherlands Institute for Space Research is the co-principal investigator institute for the TROPOMI instrument. TROPOMI is carried by the single-payload Sentinel-5P satellite. It carries out measurements of atmospheric concentrations of the relevant species for research into air quality, climate and the ozone layer. The high spatial resolution and daily global coverage make TROPOMI an important instrument in climate change research [6]. The data that TROPOMI provides allows for the detection of enhancements in concentrations that are linked to emissions, for example from iron/steel plants for carbon monoxide and gas leaks for methane.

For high accuracy emission analysis and quantification, generally a Chemical Transport Model (CTM) is used. These models simulate the atmosphere and generate concentrations of gases consistent with a given distribution of emissions. CTMs are very complex and require extensive resources in terms of time and computational load [7]. Therefore, methods have been developed that are CTM-independent. One of these methods is developed by Beirle et al. [1]. This method uses mathematical methods to estimate emissions. The method approximates transport fluxes by using reanalysis

wind fields in the continuity equation. Iterative peak-fitting is then used to distinguish and estimate emissions from large point sources [1]. The main advantage of this method is the ability to identify and separate closely-spaced point sources. The second advantage is the ability to estimate a full grid of emissions. This method is developed for the short-lived pollutant nitrogen dioxide. However, no such model exists yet for carbon monoxide and methane. Therefore, this research will investigate if the model by Beirle et al. [1] can be adapted for carbon monoxide and methane. To construct the models for CO and methane, the different characteristics of the gases need to be taken into consideration.

The main goal of this research is to test the CTM-independent model by Beirle et al. [1] for carbon monoxide and methane. This adaptation is not straightforward as both the observations (e.g. worse TROPOMI coverage for both CO and methane) and characteristics of the gases (longer lifetime in the atmosphere, different spatial and temporal patterns of emissions) differ greatly. The CO model is developed and fully optimized, while the model for methane is developed, but the optimization of the model is not within the scope of this research. To test the validity of the model, the outcome of the mathematical model are compared with the emission estimates based on chemical transport model simulations and emission inventories.

1.2. The Beirle Method

The Beirle method is a method named after Beirle et al. [1]. The method uses daily TROPOMI NO₂ data and the temporal mean of the divergence of the flux to perform long-term emission analysis, and is especially useful in localizing and quantifying emissions from point sources (i.e. industry). In Figure 1.1, the emission profiles (E) from an urban source and a point source (industrial plant) are shown. Urban emissions tend to consist of a large area with diffuse emission sources (cars, households), while point sources have a more concentrated area of emissions.

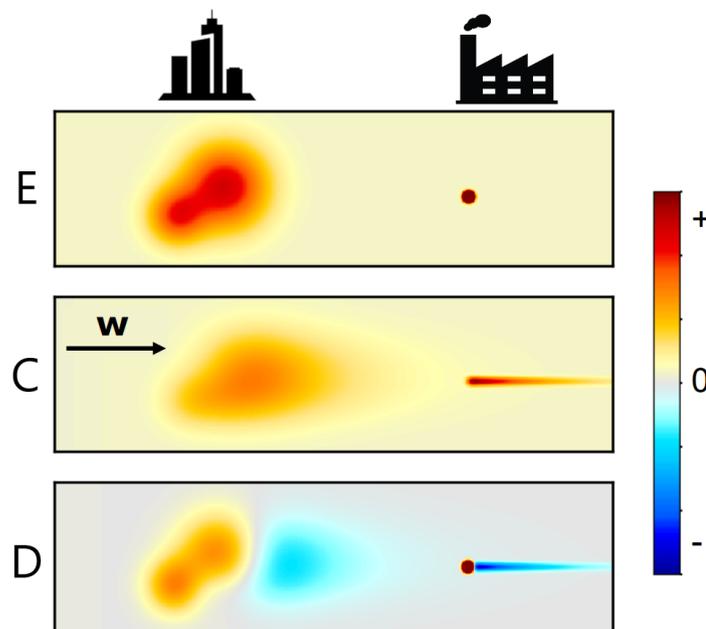


Figure 1.1: Schematic illustration for the divergence of flux profiles. The figure shows emission profiles (E) for urban and industry sources. The concentration profile (C) is slightly distorted due to the wind vector (w). The divergence (D) preserves strong flux gradients around point sources [8].

When observing the atmospheric concentration due to these emissions, the urban source shows a spatially large enhancement, slightly distorted due to the wind field (w). The point source, however, shows a plume-like enhancement down-wind of the source, parallel to the wind direction. The flux of this concentration is defined as the flow of gas particles (g m s^{-1}) over a horizontal area. The divergence of the flux (D) is highly sensitive to point sources like power plants, where spatial gradients in the flux are particularly high [8]. The divergence of urban sources consists of a large positive enhancement, followed by a large negative enhancement. The divergence of the point source sees a strong enhancement, concentrated around the area of emission (E), followed by a negative divergence along the plume. Since the divergence enhancements for point sources are spatially independent of wind direction, a long term analysis can be carried out to localize and quantify these sources.

In Figure 1.2, an example of the model result is shown. In Figure A, the TROPOMI data for a single day can be seen. Multiple NO_2 gas plumes are observed that move with the wind (blue arrows). When these concentration measurements are averaged over time (Figure B), the patterns are smeared out due to changing wind conditions (direction and speed), and the benefit of TROPOMI's high spatial resolution is lost. Due to this smearing out, long-term localization and quantification analyses using the TROPOMI concentration are infeasible. This smearing out can be prevented by averaging the divergence of the horizontal fluxes, rather than the tropospheric columns. Analysing the mean of the divergence of the flux enables the identification of point sources like Power Plants (PPs) and Cement Plants (CPs), as shown in Figure C.

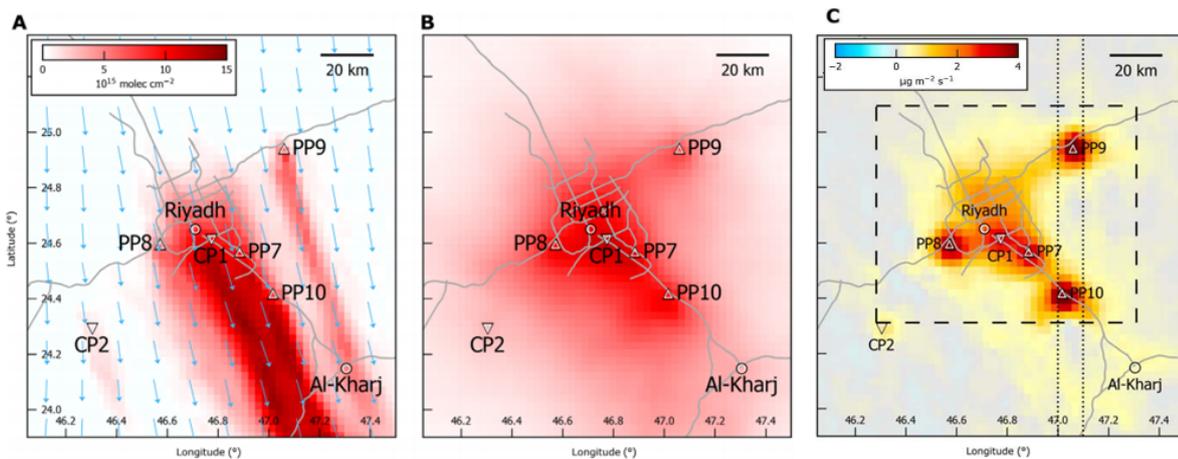


Figure 1.2: Tropospheric NO_2 column over Riyadh as derived from TROPOMI. (A) Single overpass on 17 December 2017. Arrows indicate wind vectors. (B) Temporal average for December 2017 to October 2018. (C) Emission map of TROPOMI for December 2017 to October 2018. Clear individual sources can be seen [1].

To identify emission sources from the temporal mean of the divergence, an iterative Gaussian peak-fitting algorithm is used. This algorithm consists of three steps. (1) For a given map of the temporal mean obtained from the divergence calculation, the location with the maximum divergence value in the domain is identified. (2) Around this location, a 2D Gaussian is fitted on the data. (3) The fitted peak is subtracted from the divergence map, after which the residual is used for the next iteration. This process is repeated until no more sources are identified. The Gaussian peak-fitting algorithm can be used to separate closely-spaced emission sources and to quantify a full grid of emissions.

1.3. Introduction to the Report

In [chapter 2](#), the theoretical background to the problem is given. It describes the three trace gases (NO₂, CO and CH₄). It covers the properties of the trace gases, and how these gases behave in the atmosphere. Next, the sources and sinks are described for all gases. Different techniques for atmospheric trace gases measurements are described, after which the TROPOMI instrument is elaborated in more detail. The theoretical background ends with an extensive description of the Sentinel-5P mission and the TROPOMI instrument.

The methodology for the CTM-independent emission localization and quantification method is described in [chapter 4](#). The chapter starts with a description of the TROPOMI data product and the necessary data preparation procedures to ensure the data is suitable for the divergence method. It covers the various model corrections and filters used to improve the output of the model. The chapters are concluded with a description of the pseudo data that is used to test the performance of the model.

The results of the method are described in [chapter 5](#). It starts with the results from the pseudo data. It describes the various sensitivity analyses that are used to find the optimal parameters for the CO model. After the model settings are optimised, six case studies are analysed using the optimised CO model to test the performance for real-life cases. To test the functionality of the methane model, one additional case study is described in this chapter. The report is finalised with the conclusions ([chapter 6](#)) and recommendations ([chapter 7](#)).

2

Theoretical Background

This chapter describes the three trace gases that are considered in this thesis research. It starts with an overview of Earth's atmosphere, after which the trace gas properties are elaborated. The sources and sinks are described after the first introduction to the gases. The chapter is focused on the comparison between the gases, and how different characteristics lead to other approaches in remote-sensing analysis. The TROPOMI instrument will be extensively described, after which emission quantification methods for TROPOMI will be elaborated.

2.1. Gases in the Earth's Atmosphere

The atmosphere of the Earth is a thin layer of air which is retained by Earth's gravity. The main components are nitrogen (N₂, 78.08 %), oxygen (O₂, 20.95 %) and argon (Ar, 0.934 %). Furthermore, numerous trace gases are present in the atmosphere, which are gases that make up only small fractions of the total volume. The distribution across the atmosphere is not homogeneous and varies with height [2]. Usually, the concentration of gases is given in either parts-per-million (ppm) or parts-per-billion (ppb).

The atmosphere is divided into five layers, known as (1) the troposphere, (2) the stratosphere, (3) the mesosphere, (4) the thermosphere and (5) the exosphere. The first four layers are schematically shown in [Figure 2.1](#). As can be seen in the figure, the pressure decreases almost linearly throughout these layers in the vertical direction. The transitional boundaries of the layers are defined by the reversing heat gradient throughout the atmosphere.

First, the troposphere is defined as the layer of air between the surface of the Earth up to 12-16 kilometers in height. The air in the troposphere is heated from the surface and therefore has a negative temperature gradient of -6 K km^{-1} . About 80 percent of the total atmospheric mass is present in the troposphere. Most human activity takes place in the troposphere, which results in most anthropogenic emissions being released into this layer. The second layer of air, when looking from the surface of the Earth, is the stratosphere, which stretches from approximately 12 kilometers to 50 kilometers in height. This layer of air has the absorption of solar radiation between 200 - 300 nm by atmospheric ozone as main source of heat. Therefore, the heat gradient is reversed with respect to the troposphere. In this layer, most of the Earth's ozone is present, which is commonly referred to as the ozone layer. Next, the mesosphere is reached. In this region, the temperature gradient is once again reversed due to cooling by radiative emissions from CO₂. The limit of this layer is 85 km. The stratosphere and mesosphere together are often referred to as the middle atmosphere. The fourth layer is defined as the thermosphere. This layer starts at the mesopause and, at a height of 600 km, reaches the exosphere. The thermosphere has a positive temperature gradient due to the

absorption of high-energy Ultra-Violet (UV)-radiation. Lastly, the exosphere is the final layer of air. This region has no clear upper boundary, but gradually fades into space [9, 10].

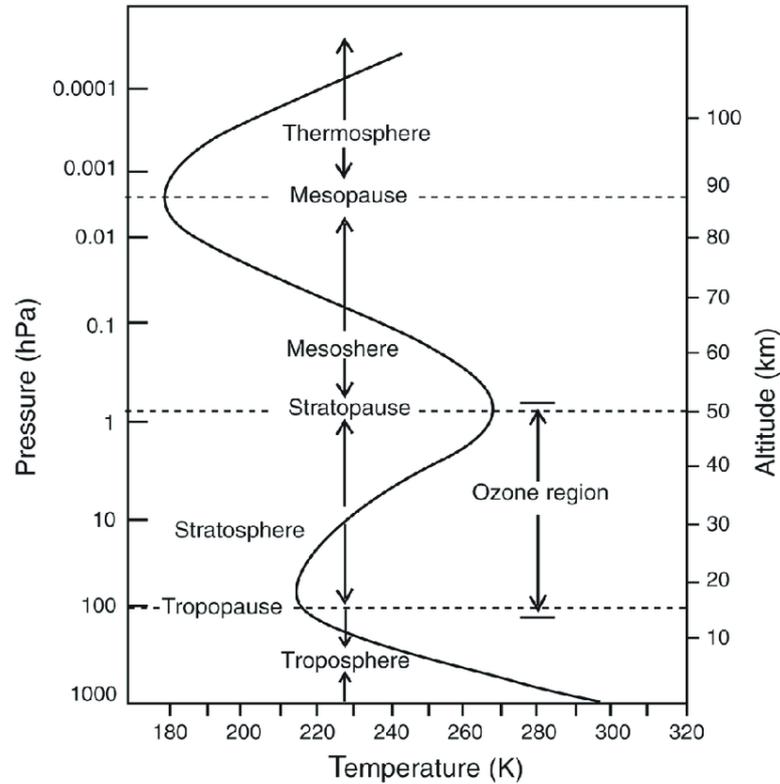


Figure 2.1: The layers of the Earth's atmosphere, classified by their vertical temperature gradient [9].

The five layers of air that together are defined as the atmosphere, are made up of different gases. Many of the trace gases that are present in the atmosphere play an important role in the regulation of the conditions on Earth, one of which is global temperature. The temperature on Earth is influenced greatly by Greenhouse Gases (GHGs). These gases, principally water vapour (H₂O), carbon dioxide (CO₂), methane (CH₄) and ozone (O₃), are naturally present in the atmosphere. These gases trap thermal infrared radiation emitted by the Earth's surface. Without these GHGs, the average surface of the Earth would be around 30 degrees Kelvin cooler [11]. This research will further investigate the emissions of NO₂, CO and methane in the atmosphere.

2.1.1. Nitrogen Dioxide

Nitrogen dioxide (NO₂) is one of seven oxides of nitrogen and is mostly generated through combustion, effectively the use of energy. It has a molecular weight of 46.01 g/mol, a melting point of 261.95 Kelvin and a boiling point of 294.3 Kelvin [12]. As previously discussed, the temperature of air in the atmosphere is often below the melting point of NO₂. However, due to the low partial pressure of NO₂ in the atmosphere (908 mmHg at 298.15 Kelvin), condensation does not occur. The lack of condensation makes that NO₂ can exist in the atmosphere in the gaseous form [13]. Nitrogen dioxide is formed by the rapid reaction between NO and O₃ (Equation 2.1), which in turn photolyzes back to NO (Equation 2.2) [14]. In this equation, $h\nu$ refers to the light photon which is absorbed by the molecule. Since these reactions are catalysed by the energy from sunlight, the NO₂/NO ratio is highly influenced by weather [15].





The O-atom from Equation 2.2 then reacts with oxygen to produce ozone. NO₂ is not a direct greenhouse gas, but acts as a precursor for the greenhouse gas ozone. This cycle of NO and NO₂ reactions takes about one minute during daytime. Since it has no net effect on ozone, it is referred to as a null cycle.

Nitrogen dioxide is also produced by the oxidation of nitric oxide by oxygen. This mainly takes place at high temperatures during the combustion process. As the temperature of the combustion process increases, the percentage of emitted nitrogen dioxide is decreased. Actually, for normal combustion, only 5 to 10 percent of the nitrogen oxides are emitted as nitrogen dioxide. The remaining 90 to 95 percent is emitted as nitric oxide. This does, however, strongly vary with source type [16]. Nitrogen dioxide has an absorption spectrum that lies in the UV and visible spectrum, between approximately 350 and 600 nm [6]. Due to the fast cycle of NO and NO₂ which takes only a minute during daytime, both gases will be referred to as NO_x from this point forward.

The main sink of NO_x is the oxidation to HNO₃ in the daytime, according to Equation 2.3, where M denotes a non-reacting molecule (chaperone) that absorbs a portion of the energy to ensure a stable reaction product [2]. At night, NO₂ is oxidized by O₃, which forms N₂O₅ by conversion of the NO₃ radical. This formation can only take place at night, since during the daytime NO₃ is photolyzed back to NO₂ within a few seconds. The products of the oxidation of NO_x, which are HNO₃ and N₂O₅, are eventually converted back to NO_x and are serving as reservoirs. However, stratospheric N₂O₅ can also be converted to HNO₃ in the presence of aqueous aerosols [17]. The combination of NO_x and its reservoirs is referred to as NO_y. The main importance of nitrogen oxides (NO_x) is the role it has in the regulation of the Earth's ozone distribution [14].



The global mean concentration of NO_x in the troposphere is estimated as 1.7 ppb [18]. This value is also called the mixing ratio or mole fraction, which is the number of moles for a specific gas per mole of air [2]. It is given in mol/mol or as unit of volume v/v. Mixing ratios of trace gases are usually given in ppm or ppb. However, the local mean mixing ratio of NO_x varies greatly per region. For instance, urban NO_x concentrations are observed to be much higher than the global average. The urban NO_x concentration in Europe saw a peak in the early 1990s of 30 ppb, but has decreased from 28 ppb in the late 1990s to 11 ppb present day. For roadside areas, it is observed to have decreased from 32 ppb to 18 ppb for the same time period. For rural background, the decrease has been smaller, with values of around 10 ppb in the late 1990s to 4 ppb present day [19]. This decrease in NO_x can be explained due to stricter emission standards for new road vehicles. Also, the power generation relies less on the usage of coal, especially in first-world countries. In recent years, diesel engines have become less popular [20]. Since diesel-fuelled cars emit more NO_x, this could explain the decrease of NO_x emission in roadside areas. The reduction in rural background concentrations could be explained due to the overall decrease in NO_x emissions across Europe.

There exists a daily variation of weekday mean concentration of NO_x in the troposphere, as observed in the UK in 2019 and shown in Figure 2.2. This can be explained by the lower number of vehicle hours in the weekends. The mean concentration of NO₂ is observed to be up to 22% lower in the weekends in the UK. Hourly variations for roadside areas tends to be related to rush hour. Peaks can be observed around 9 AM and 6 PM, with a daily minimum around 3 AM. These hourly means are similar for urban areas, but with a peak at 9 PM compared to the 6PM for roadside. The rural areas tend to have a smaller hourly variation, and barely noticeable minima and maxima [19].

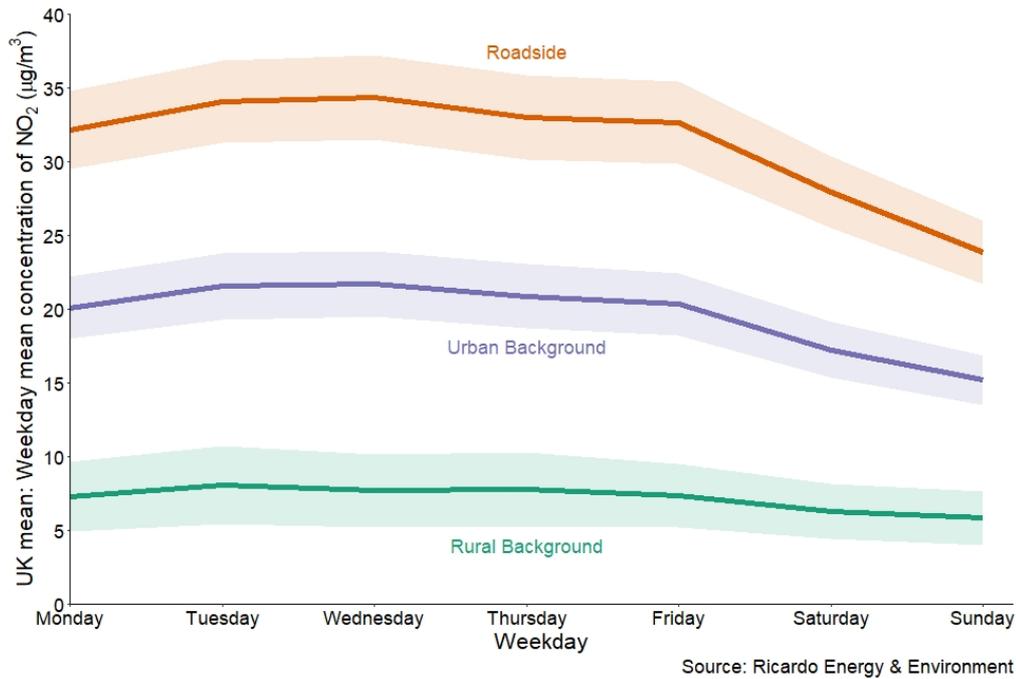


Figure 2.2: Daily NO₂ concentrations as observed by UK ground stations for roadside, urban and rural areas. 1 ppb \approx 0.53 $\mu\text{g m}^{-3}$ [19].

2.1.2. Carbon Monoxide

Carbon monoxide (CO) is a toxic, colorless and odorless gas. It is an important atmospheric constituent when considering air quality [21]. The production of CO is related to the partial oxidation of carbon-containing compounds. This occurs when there is not enough oxygen in the combustion process to produce carbon dioxide (CO₂), or is achieved due to low mixing of air and fuel [22]. The reaction is shown in Equation 2.4.



CO is not poisonous, but it does have a negative temporary effect on the human respiratory system. CO can attach itself to hemoglobin in the human red blood cells, which in turn prevents the uptake of oxygen [23]. CO has a molecular mass of 28.0 g/mol, which makes it slightly lighter than air (\approx 28.8 g/mol). It has a melting point of 68 Kelvin and a boiling point of 82 Kelvin. This results in CO being present in the atmosphere in the gaseous phase. CO itself is not a strong greenhouse gas, but it does influence radiative forcing indirectly by affecting concentrations of greenhouse gases, such as methane and ozone. Therefore, it does not cause climate change directly, but greatly influences it [24]. Carbon monoxide reacts with the hydroxyl radical (OH) which triggers a chain of chemical reactions, ultimately producing carbon dioxide, NO_x and the hydroxyl radical. Also, since OH is the most important sink of methane, CO has an indirect influence on the concentration of methane. Furthermore, CO is the largest principal sink of tropospheric OH, with up to 40 percent of the OH radicals reacting with carbon monoxide in today's troposphere [25, 21]. The hydroxyl radical is in turn important since it removes most climate-damaging trace gases by oxidation. This is referred to as the oxidation capacity of the atmosphere and without this process, climate and atmospheric composition would be very different from what it is today [26]. The chain of reactions that is initialized by CO and ultimately leads to the formation of ozone is shown in Equation 2.5.



Carbon monoxide has an absorption spectrum that lies in the infrared, with two peaks at approximately $1.6 \mu\text{m}$ and $2.3 \mu\text{m}$ in the Short-Wave InfraRed (SWIR) and $4.6 \mu\text{m}$ in the Medium-Wave InfraRed (MWIR) [27].

The concentration of carbon monoxide in the atmosphere is generally around 100 ppb. Higher levels are observed in urban areas, compared to rural areas [28]. General background concentrations of carbon monoxide on a global scale are found to be between 50 and 120 ppb. The concentration of CO in urban regions, however, depends highly on weather conditions and topography, as well as the density of combustion-powered vehicles and the distance from the traffic [23]. The concentration of CO has increased gradually from around 90 ppb in the 1850s to 110 ppb in the 1950s. From the early 1600s up to the 1850, CO levels were at a reasonably constant 90 ppb [25]. From the beginning of 2000, CO levels in the troposphere have decreased from 125 ppb in 2000 to 105 in 2014, as shown in Figure 2.3. This change in the global CO concentration can be explained by technological innovation. Vehicles and industries are, due to technological advancements, polluting less compared to previous times. The peak around 2003 can be explained by large scale forest fires in Russia. The dip after 2008 can be explained by the global financial crisis, which decreased global manufacturing activity [24].

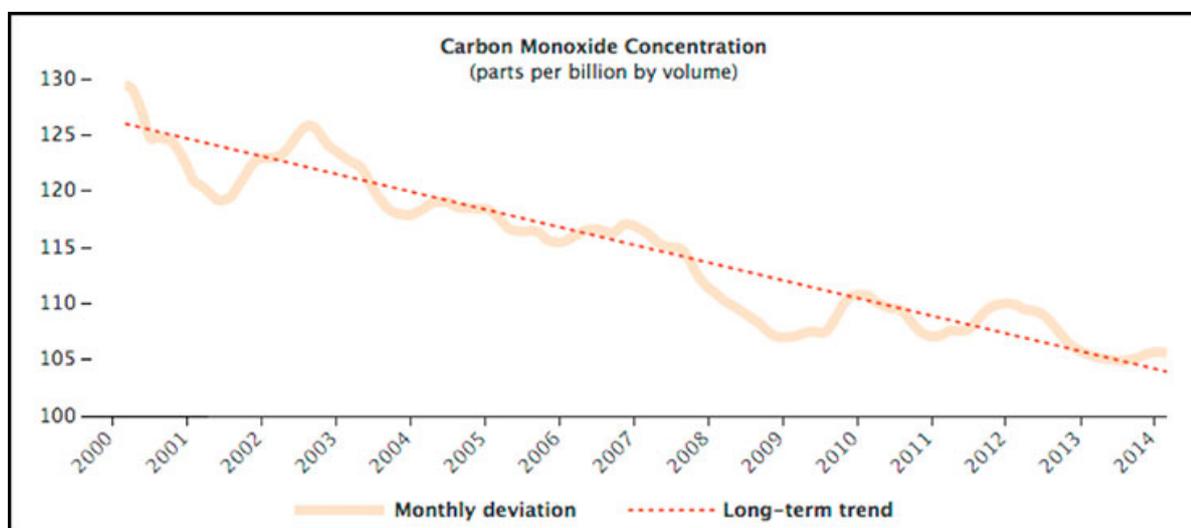


Figure 2.3: Global mean carbon monoxide levels at an altitude of 3,700 meters, as measured by the MOPITT instrument [24].

2.1.3. Methane

Methane (CH_4) is a colorless and odorless gas and is the most abundant organic trace gas in the atmosphere. Methane is third in line as most abundant greenhouse gas in the troposphere, following water vapour and carbon dioxide (CO_2) [29]. It has a molar mass of $16.043 \text{ g mol}^{-1}$, a melting point of 90.7 Kelvin and a boiling point of 111.6 Kelvin, which makes that it is present in the atmosphere in the gaseous form. Most methane that is present on Earth is produced by microbes which are present in landfills and other soils, cattle, insects like termites, sediments and rice fields [30]. The reaction which produces methane by microbes is shown in Equation 2.6.



Methane can also be produced geologically due to the break-up of organic matter at high pressure and temperature. The reaction in Equation 2.6 is also seen in industrial processes. However, there is little need for the production of methane for industrial purposes. Methane is slowly oxidized by

photo-chemical reactions in the atmosphere which ultimately produces ozone and carbon dioxide. Methane can also be oxidized by the hydroxyl radical to produce, among other gases, CO and ozone, if there is sufficient NO_x present [29]. Methane, just as CO, controls the amount of OH in the troposphere.

The relative influence of different trace gases on global warming is defined as the Global Warming Potential (GWP). This is defined as the heat that is absorbed by a GHG, relative to the heat that would be absorbed by the same emitted mass of carbon dioxide over a certain time period. Therefore, the GWP of CO_2 is set to one [31]. The GWP also includes indirect effects on global warming. For instance, methane has a direct influence on global warming, but also acts as a precursor for atmospheric ozone. Gases with a higher GWP have a higher impact on global warming compared to CO_2 over the same time period [32]. These GWPs are defined for the short-term of 20 years (GWP_{20}) and the long-term of 100 years (GWP_{100}). Methane has a GWP_{20} of 82.5 and a GWP_{100} of 29.8 [33]. The short-term GWP for methane is much larger than the GWP of carbon dioxide for the same time-span. Carbon dioxide has a very long lifetime in the atmosphere (up to 1000 years [34]), which results in CO_2 emissions causing increased CO_2 concentrations for over a thousand years. The lifetime of carbon dioxide is not straightforward, since CO_2 is not destroyed over time, but instead moves through different parts of the ocean-atmosphere-land system. To simplify this value, often a lifetime of 100 years is chosen. The lifetime of methane, for instance, is much shorter, as will be elaborated in subsection 2.1.5. However, a molecule of methane can absorb substantially more energy than a molecule of carbon dioxide, which makes that methane has a higher GWP than carbon dioxide [32]. This results in methane being an excellent target for climate change mitigation research.

The mean global methane concentration in the atmosphere has increased by 0.5-0.8% annually since the industrial revolution [35]. In the 1750s, the global mean methane concentration was estimated to be around 722 ppb. This rose to 1890 ppb in 2021. The global monthly mean tends to vary slightly, as can be seen in Figure 2.4.

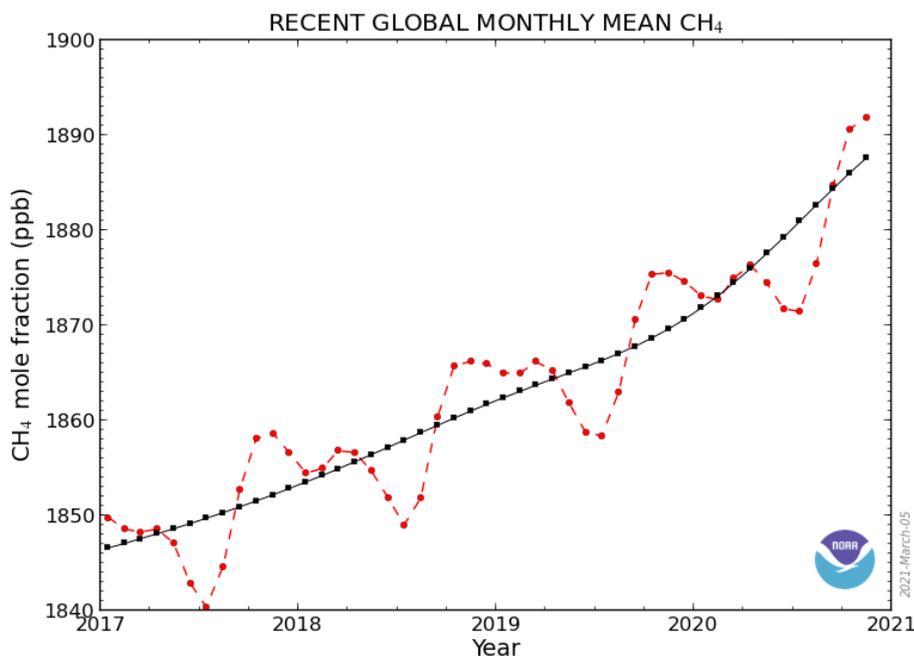


Figure 2.4: Globally averaged, monthly mean atmospheric methane abundance determined from marine surface sites, as published by the Global Monitoring Division of NOAA's Earth System Laboratory [36].

These monthly variations are due to the seasonality of the OH sinks and sources, like wetlands, rice fields and the burning of biomass. Maximum monthly concentrations are observed around September and October with a minimum in the summer months [37]. This is due to the maximum concentration of OH being in the summer. Furthermore, this figure also shows the rapid increase of mean methane levels in the Earth's atmosphere in the last four years. Since emissions, and therefore concentrations, of methane are rapidly increasing, methane is the second most important human-induced greenhouse gas, with carbon dioxide being the first [38]. Methane is responsible for more than 18 percent of the heat that is generated from greenhouse gas concentrations in 2019 [37].

2.1.4. Mixing Ratio

In the previous sections, the mixing ratios for NO_x, CO and methane are given. To more easily compare the values, the mixing ratios of the trace gases are shown in Table 2.1 [18]. The values given in the table are global mean values. The mixing ratios of trace gases differ highly per location. For instance, the local mixing ratio for NO_x of the Asia-Pacific region is 4.7 ppb, while in the more remote regions like Oceania it is observed to be < 0.1 ppb [18]. For carbon monoxide, the global mean concentration varies between 50 - 120 ppb [23]. For methane, the mixing ratio is seasonably varying, but the global mean is observed to be around 1890 ppb. However, it is rapidly increasing [36]. The global mean concentration of trace gases is highly dependent on the lifetime of the specific trace gas. The longer the lifetime, the more it mixes in the atmosphere. Gases with short lifetimes are not transported over large distances and are not well-mixed and therefore have a lower mixing ratio.

Table 2.1: Global means for selected trace gases.

	Mixing ratio (mol/mol)	Mixing ratio (ppb)
Nitrogen oxides (NO _x)	1.7×10^{-9}	1.7
Carbon monoxide (CO)	1×10^{-7}	100
Methane (CH ₄)	1.89×10^{-6}	1890

2.1.5. Lifetime

The influence of greenhouse gases on climate change is largely determined by their lifetime in the atmosphere. The lifetime of a trace gas is calculated by using the global burden (B) and the global integrated loss (L). The global burden is the number of molecules of a trace gas (n) integrated over the volume (V) of the atmosphere [39]:

$$B = \int n \cdot dV \quad (2.7)$$

The global integrated loss (L_{GIL}) is the local sinks (l) in molecules per unit volume per unit time, integrated over the volume. The local sink factor (l), if constant, can also be written as $l = \alpha n$, where α is the local loss frequency. If a trace gas has no local sinks, then $\alpha = 0$. The equation for the global integrated loss can now be written as:

$$L_{GIL} = \int l \cdot dV = \int \alpha n \cdot dV \quad (2.8)$$

The equation for the global lifetime (τ_g) can now be written as:

$$\tau_g = \frac{B}{L_{GIL}} \quad (2.9)$$

As can be seen from Equation 2.9, the lifetime of a trace gas is not only dependent on the distribution and magnitude of the loss frequency, but also on the spatial distribution. These factors again

depend on atmospheric transport, loss rates and location of the emissions. As these factors change over time, the lifetime in turn also changes. This makes that there is no unique lifetime for specific trace gases [39].

The lifetime of NO_x relates non-linearly to its own concentration [40]. This is mainly due to NO_x being a precursor its own sinks. In summer months, it decreases to 3 hours, while in the winter months it can get up to 13 hours [41]. Kenagy et al. [42] found the lifetime of NO_x in the winter during the day to be up to 29 hours, as opposed to a lifetime of 6.3 hours during the night. This is mainly due to lower mean temperatures in wintertime [42]. However, recent studies define the mean lifetime of NO_x as 4 hours [43]. The short lifetime of NO_x makes that the determining of NO_x emissions from observed NO_x columns is relatively straightforward [44]. Carbon monoxide has a lifetime of about 1 - 3 months [45]. This time period is long enough for CO to be transported over long distances by winds, but not long enough for the gas to be mixed evenly throughout the atmosphere [24]. Methane has a lifetime in the atmosphere of 9.1 [8.2 - 10.0] years [46]. This duration is sufficient for methane to be transported over long distances and be reasonably well-mixed throughout the atmosphere. A total overview of the lifetimes for NO_x , CO and CH_4 can be found in Table 2.2.

Table 2.2: The different lifetimes for selected trace gases.

	Lifetime
Nitrogen oxides (NO_x)	3 - 29 hours
Carbon monoxide (CO)	1 - 3 months
Methane (CH_4)	8.2 - 10 years

2.2. Sources and Sinks

In the previous section, sources and sinks of NO_x , CO and methane are elaborated. This section aims to quantify those sources and sinks. It is well known that CO and NO_x have significant anthropogenic and natural sources. Their budgets are taken from, respectively, Stavrou et al. [47] and Holloway, Levy, and Kasibhatla [48]. The source and sink budgets of methane are taken from Saunio et al. [49].

2.2.1. Nitrogen Dioxide

NO_x emissions are dominated by anthropogenic sources. These source types make up around 65% of total global NO_x emissions [50]. The main anthropogenic source is fossil fuel combustion. About 50% of the fossil fuel emissions are due to stationary sources like power plants, cement plants and industrial boilers. The other half is found to be from motor vehicles [51]. Natural sources of atmospheric gases are all source types that are not directly caused by humans. For NO_x , natural sources make up about 35% of total NO_x emissions globally. The main natural source is emission from microbial processes in soils, which makes up around 17% of the total NO_x emissions globally [50]. However, soils can also be identified as an (partly) anthropogenic source due to fertilizer applications. Fires are estimated to contribute around 9% to global NO_x emissions. Furthermore, lightning emissions contributes an average of 6% to global yearly NO_x emissions [52]. The remainder of NO_x emissions comes from oxidation of ammonia and volcanic activity [51]. However, the uncertainties are significant for natural sources [47]. In Table 2.3, the major NO_x sources in Teragram ($\text{Tg} = 10^{12}$ g) are shown.

The main sink of NO_x is the reaction with the hydroxyl radical to produce nitric acid (HNO_3). This is a daytime reaction and takes place 5-10 times faster in summertime than in the wintertime [51]. The global budget that is the result of this reaction is difficult to define, since it is hard to measure the rate constant for the thermolecular association reaction [47]. Hydrolysis of N_2O_5 at the surface

Table 2.3: Global annual NO_x emissions [47].

Sources	Global emissions [Tg a ⁻¹]
Anthropogenic	28.7
Fires	5.2
Soil	14.7
Lightning	4.6
Total source	53.2

of aerosols is an important NO_x loss process during nighttime. Its impact is largest in high-polluted areas in wintertime, since nights are longer compared to summertime. Other sinks of NO_x include dry deposition to vegetation and soils, which is a direct sink of NO_x. An indirect sink is found in the wet and dry deposition of organic nitrates. In Table 2.4 an overview of the sinks is given. From a comparison with the sources overview, it can be seen that there is a global annual NO_x emission surplus of 0.5 Tg.

Table 2.4: Global annual sinks for NO_x [47].

Sinks	Global sinks [Tg a ⁻¹]
NO _x + OH	30.1
Aerosols	9.7
Other	12.9
Total sink	52.7

2.2.2. Carbon Monoxide

For CO, the emissions are found to be 60% anthropogenic and 40% natural. Anthropogenic emissions are mainly the result of incomplete combustion of carbonaceous materials. Global annual CO emissions in 2000 were found to be 2491 ± 361 Tg. As the global CO concentration since 2000 has slightly decreased, it can be assumed that this is the results of decreased CO emissions. Therefore, the values described in this section need to be scaled to this decrease. The main anthropogenic source of CO emissions is biomass burning, which contributes 748 Tg/a (498 Tg/a CO in the Northern Hemisphere, 250 Tg/a CO in the Southern Hemisphere). This includes burning of savanna, forests, agricultural residue, fuelwood, and animal waste. It does have to be said, however, that not all forest fires are anthropogenic, and should be classified as natural sources. The second largest anthropogenic source is fossil fuel combustion with a total of 300 Tg/a. From this emission value, 282 Tg is emitted in the Northern hemisphere and 18 Tg in the Southern hemisphere [48].

The chemical/natural sources of CO are dominated by two oxidation types: biogenic HydroCarbon (HC) oxidation and methane oxidation. The oxidation of HCs results in 683 Tg/a globally. As opposed to anthropogenic sources, there is a minimal difference in emissions for the Northern and Southern hemisphere. Lastly, the reaction of methane with the hydroxyl radical yields CO as a product. The production of CO is therefore highly dependent on the distribution of CH₄ and OH. As methane concentrations are slightly higher in the Northern hemisphere as opposed to the Southern hemisphere, there is a minimal surplus of CO produced in the Northern hemisphere. The global value results in 760 Tg annually [48]. An overview of the values can be found in Table 2.5. Other minor CO sources are vegetation and oceans.

Table 2.5: Global annual CO sources [48].

Sources	Global emissions [Tg a ⁻¹]
Fossil fuel	300
Biomass burning	748
Biogenic HC oxidation	683
Methane oxidation	760
Total source	2491

CO is mainly removed from the atmosphere by tropospheric oxidation with the hydroxyl radical to form carbon dioxide. As can be seen in Table 2.6, there is a significant uncertainty in the sink budget from OH oxidation [48]. Other sinks of CO include chemical reactions in the stratosphere and soil uptake. When the sinks of CO are compared to the source, it can be seen that it is not quite clear whether there is an CO emission surplus or not. When the global mean is taken (CO sinks: 2491 Tg/a), it is clear that the global CO concentration is decreasing.

Table 2.6: Global annual sinks for carbon monoxide [48].

Sinks	Global sinks [Tg a ⁻¹]
Oxidation by OH	1400 - 2600
Stratosphere	100
Soil uptake	250 - 640
Total sink	2100 - 3000

2.2.3. Methane

As opposed to NO_x and carbon monoxide, the global concentration of methane has continued to increase in the last years [36]. Methane emissions can be grouped into three categories: biogenic, thermogenic and pyrogenic. Biogenic sources contain methane-producing microbes (methanogens). Examples of biogenic methane sources are wetlands, rice paddies, agriculture and waste. Thermogenic methane is formed over millions of years through geological processes, from which the emissions mainly consist of exploitation of coal, oil and natural gas. Lastly, pyrogenic methane is produced by the incomplete combustion of biomass and soil carbon, mainly during biomass burning and combustion of fossil fuels [53].

The methane budgets that are shown in this chapter are all taken from research by Saunio et al. [49]. The main anthropogenic source of methane is agriculture and waste. This includes cattle, rice cultivation and landfills. This source has an estimated total emission value of 227 [205-246] Tg/a in 2017. When looking at current emissions, this value needs to be scaled up. The scaling factor will not be identical for every emission source, but differs per source type. The second largest anthropogenic source is fossil fuel exploitation, averaging around 108 [91-121] Tg/a. This includes coal plants and gas transport pipelines, as well as compressor stations. This is followed by biomass burning, averaging around 28 [25-32] Tg/a.

For natural methane sources, wetlands are found to be the largest at 194 [155-217] Tg/a. Furthermore, there are many minor sources that contribute to global methane emissions. Among others, these include fresh water lakes, wildfires, wild animals, oceans and permafrost. Combined, they result in 39 [21-50] Tg emitted annually [49]. An overview of emissions is given in Table 2.7.

Table 2.7: Bottom-up global annual sources for methane [49].

Source	Global emissions [Tg a ⁻¹]
Agriculture and waste	227 [205-246]
Fossil fuels	108 [91-121]
Biomass burning	28 [25-32]
Natural wetlands	194 [155-217]
Other natural sources	39 [21-50]
Total natural sources	232 [194-267]
Total anthropogenic sources	364 [340-381]
Total methane sources	596 [572-614]

The methane sinks are dominated by oxidation with the hydroxyl radical. This oxidation occurs mainly in the troposphere and accounts for about 80% of the global methane sink. Other large sinks are, among others, oxidation by bacteria in aerated soils (4%), oxidation by chlorine radicals and oxygen in the stratosphere (3%) and reactions with chlorine radicals from sea salt (3%). Furthermore, the soil sink makes up 6% of the total methane sinks [53]. An overview can be found in Table 2.8. From the source and sinks, it can be seen that there is a global annual methane imbalance of 25 Tg in 2017, with a large uncertainty. This surplus has resulted in the increase of the global mean concentration of methane over the last years.

Table 2.8: Global annual sinks for methane [49].

Sink	Global sink [Tg a ⁻¹]
Soil uptake	40 [37-47]
Total chemical loss	531 [502-540]
Total sink	571 [540-585]

This thesis research focuses on anthropogenic emission sources of carbon monoxide and methane. For carbon monoxide, the main sources are coal-, gas- and oil-fired power plants and other industrial processes, which include chemical production, petroleum refining and metals production. For methane, the main anthropogenic industrial processes can be traced back to fossil fuel exploitation such as coal mines and operational oil and gas facilities. Furthermore, landfills are a major factor amongst anthropogenic methane sources.

2.3. Measuring Atmospheric Trace Gases

There are different methods to measure atmospheric trace gas concentrations, ranging from static in-situ instruments to remote-sensing techniques with aircraft or satellites. Since the divergence method uses large grids of trace gas concentration data, this research is focused on remote-sensing techniques. This section will describe the TROPOMI instrument, which makes use of absorption spectroscopy.

2.3.1. Physical Measurement Principles

The absorption bands of CO and methane are used to detect the concentration of the trace gases in the atmosphere. This is done using absorption spectroscopy, which is widely used in remote sensing. TROPOMI uses backscattered solar radiation for the detection of trace gases, as shown in Figure 2.5 (a). Photons from the sun pass through an atmospheric column, reflect on the Earth's surface, and are detected by the remote-sensing instrument. Each trace gas in the atmospheric col-

umn absorbs energy from the photons at specific wavelengths. The photons that are detected by the instrument can be analysed for dips in the spectrum, after which an estimation can be made of the presence of specific trace gases in the measured atmospheric column. It is important for remote-sensing instruments to pass over a specific location at the same local solar time, since consistency in viewing geometry conditions and illumination must be ensured. This way, measurements at different days are inter-comparable. To achieve this, the instrument must be carried by a satellite which is in polar sun-synchronous Low Earth Orbit (LEO) [5].

The trace gas concentrations retrieved by remote-sensing instruments are generally expressed as the *total column dry air atmospheric mixing ratio*. The dry-air mole fraction of CO is defined as the total number of CO molecules in a vertical column above a unit surface divided by the corresponding total number of molecules of dry air in that column. This is done to ensure that the measurement is mostly insensitive to varying surface altitude and pressure differences. However, due to a changing relative contribution of the stratosphere, the total column dry air atmospheric mixing ratio has a dependency on surface elevation, and therefore is not fully insensitive to surface altitude. This will be further elaborated in [chapter 4](#). This insensitivity is shown in [Figure 2.5](#) (b). The CO concentration is dependent on the atmospheric pressure, but the dry-air mole fraction is taken relative to the O₂ concentration. This levels out changes in the XCO measurement due to a varying atmospheric pressure [3].

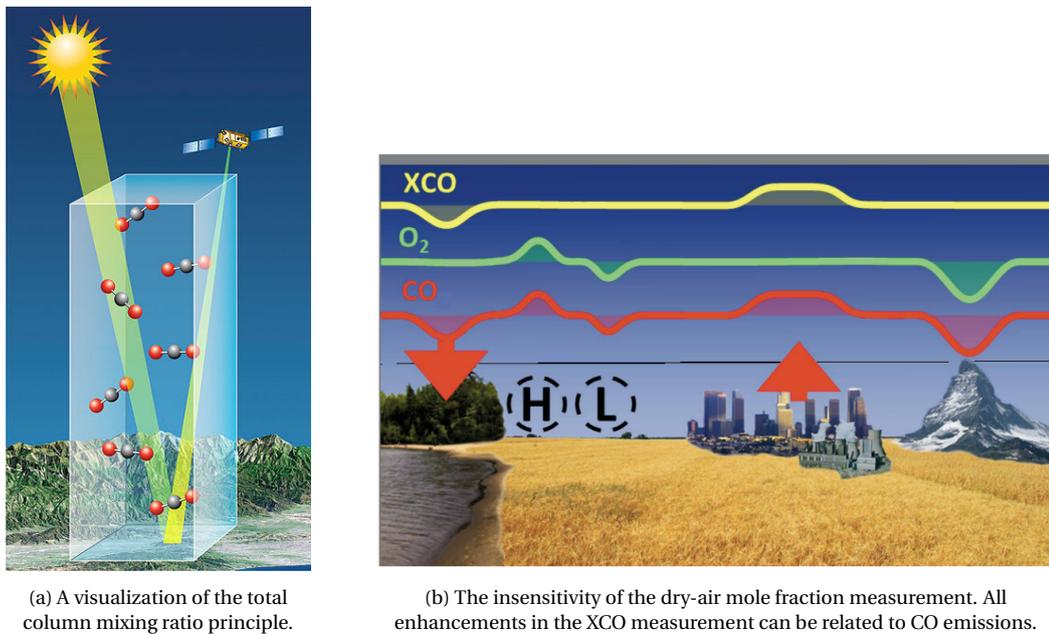


Figure 2.5: Principles used in remote-sensing measurements [3].

One last vital detail to note is that remote-sensing instruments do not measure emissions, but concentrations. Since the instrument's measurements always include the tropospheric background, emissions cannot be easily isolated. This makes that methane emission detection is more difficult than CO emission detection due to the higher background values of methane compared to CO, which in turn is more difficult than NO_x emission detection.

2.4. TROPOMI

The Sentinel-5 Precursor (Sentinel-5P) satellite is part of the Global Monitoring of the Environment (GMES) project by the European Community (EC). It is a Dutch initiative and is realized in cooperation with the European Space Agency (ESA). The GMES initiative is established to help Europe reach

its goals regarding global governance of the environment and sustainable development. This is done by providing high quality data, information and knowledge in a timely fashion [6]. The Sentinel-5P mission is a single-payload satellite that orbits the Earth in LEO at approximately 824 km. It has a sun-synchronous orbit with a mean overpass time of 13:30 h and a period of 101 minutes. A sun-synchronous orbit means that the overpass on any point on the surface of the Earth is at the same local solar time, which is necessary to maintain consistent lighting and sun angle. Sentinel-5P was launched in October 2017 and has TROPOMI as payload. TROPOMI provides information on concentrations of trace gases and aerosols with daily global coverage, which means the satellite has the possibility to monitor the entire surface of the Earth every day. It has a swath of 2600 km, so 14 orbits are needed to obtain daily global coverage. The measured trace gases include ozone (O_3), carbon monoxide (CO), methane (CH_4), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), formaldehyde (CH_2O), aerosols and cloud properties. These trace gases are important for research into air quality, the ozone layer and climate forcing. As the name suggests, the Sentinel-5P mission connects the previous Ozone Monitoring Instrument (OMI) and Scanning Imaging Absorption spectrometer for Atmospheric Cartography (SCIAMACHY) missions with the future Sentinel-5 missions. Sentinel-5P works in close cooperation with the NASA-NOAA's Suomi National Polar-orbiting Partnership (Suomi NPP) satellite, which includes the use of Visible/Infrared Imager and Radiometer Suite (VIIRS) for high resolution cloud information. Sentinel-5P flies within 5 minutes of Suomi NPP, trailing behind it. Suomi NPP provides cloud mask data with a higher spatial resolution than Sentinel-5P [6].

TROPOMI measures trace gas concentrations by analysing reflected sunlight from the Earth's atmosphere in four spectral bands: UV, visible, Near-Infrared (NIR) and SWIR. The measurement principle is depicted in Figure 2.6. TROPOMI measures a strip on the Earth by using a two-dimensional detector. This is done for a period of 1 second, in which the satellite moves about 7 km. This is why the spatial resolution in the latitudinal direction is 7 km. After August 2019 this resolution was increased to 5.5 km, which is achieved by carrying out a measurement every ~ 0.8 seconds. The width of the strip, or the swath, is 2600 km. Each second, a new measurement is started. Therefore, the instrument scans the Earth as it moves [6].

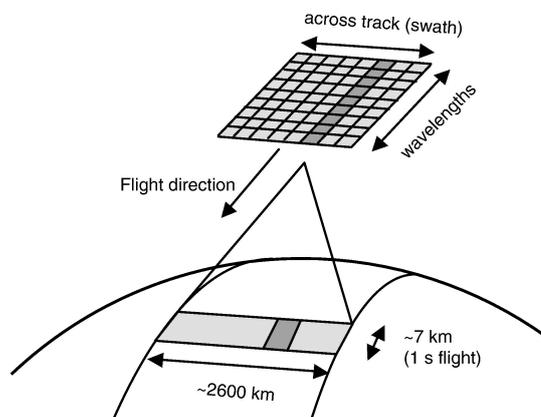


Figure 2.6: The TROPOMI measurement principle, which shows the resolution of 7 km and the swath of 2600 km. All ground pixels are measured at the same time [6].

2.4.1. TROPOMI Trace Gas Concentration Retrievals

The TROPOMI data products that will be addressed in this thesis research are shown in Table 2.9. In this table, product accuracy is defined as the mean deviation between the actual value and the measured value. The product precision is defined as the variation due to randomness, such as noise. These values are estimates and are based on retrieval simulations, as well as on findings

from previous OMI and SCIAMACHY missions. The accuracy of the methane product depends heavily on the cloud clearing. This procedure decreases the accuracy of the product, since it has to balance false cloud detection in clear scenes with cloud contamination. To minimise this error, TROPOMI uses a cloud mask from the VIIRS data that is provided by the Suomi NPP satellite, instead of using its own UVNS data [6]. In the following section(s), the TROPOMI data products will be elaborated in more detail. Additional data product information can be found on the website of the TROPOMI Mission Performance Centre: <http://www.tropomi.eu/data-products/mission-performance-centre>.

Table 2.9: TROPOMI data products for individual observations [6].

Product	Accuracy	Precision	Remarks
NO ₂			
Tropospheric column	10%	$1 \cdot 10^{15}$ mol/cm ²	Cloud-free, polluted conditions
Total column	$1 \cdot 10^{15}$ mol/cm ²	$1 \cdot 10^{15}$ mol/cm ²	Background concentration
CO			
Total column	15%	10%	
CH ₄			
Methane	2%	0.6%	Cloud free

Nitrogen Dioxide

The spatial and temporal distributions of nitrogen dioxide vary strongly due to the short lifetime. The high spatial resolution of TROPOMI is therefore very important. Variations in the NO₂ concentration can be detected more easily and emission sources can be better quantified. Since most of the NO₂ in the troposphere is present in the boundary layer, small clouds and sub-pixel cloud variability are a major source of error on spatial scales at or below TROPOMI's spatial resolution [54]. Therefore, the NO₂ data is reliable up to a specific cloud fraction [6].

Carbon Monoxide

For cloud-free and low aerosol concentration scenarios, TROPOMI SWIR measurements are experiencing little atmospheric scattering. With scattering, there is no energy transformation, but the spatial distribution of the photon's energy changes. The low scattering results in most of the measured photons being reflected at the Earth's surface. This light path causes the measurements being influenced by CO in the tropospheric column. Largest CO concentrations are measured close to emission sources in the lower troposphere. Because of the sensitivity for variations in CO concentration in the lower troposphere, TROPOMI SWIR measurements can be applied to relate satellite measurements to emission sources. There is one downside to limiting the CO observations to cloud-free scenarios only. This would limit the number of CO retrievals dramatically, since only 11% of observations are on clear sky regions [55]. However, observations with low cloud fractions have sufficient sensitivity for CO. Thus, the TROPOMI SWIR retrieval algorithm accounts for both cloud-free and partial cloud coverage scenes [21].

Methane

Atmospheric methane measurements make, just like carbon monoxide measurements, use of spectroscopic measurements of sunlight back-scattered by the Earth's surface and atmosphere in the SWIR spectral range. The path of the measured photons is largely influenced by aerosols and cloud coverage. Since reliable prior information about these parameters is generally not available, methane retrieval methods are simultaneously inferring gas concentrations and a correction for scattering effects [21]. For methane, only cloud-free measurements are used, since otherwise the accuracy of the measurement is not sufficient. The TROPOMI methane retrieval makes use of SWIR and NIR

bands to both obtain methane total column concentrations as well as atmospheric scattering properties to minimise scattering induced retrieval errors on methane. The SWIR band is used for the methane concentrations, while the SWIR-NIR combination constraints the atmospheric scattering properties. The errors induced by the residual scattering are found to be mostly below 1% [56].

2.4.2. Data Quality and Known Data Issues

SRON recommends to use the CO data with $qa > 0.5$. The qa-value parameter definition is shown in Table 2.10, which is taken from the TROPOMI CO ATBD.

Table 2.10: CO qa-value parameter definition.

qa-value	Remark
1.0	Clear-sky and clear sky like observations
0.7	Mid-level clouds
0.4	High clouds, experimental data set
0.0	Corrupted or defective data

For methane, it is recommended to use $qa > 0.5$ to avoid misinterpretation of the data [57]. The quality definition values for methane are given in Table 2.11, which is taken from the TROPOMI methane ATBD. As for CO, TROPOMI CH₄ data has some known issues as well. First of all, filtering on $qa > 0.5$ does not remove all pixels with poor quality. In some cases, pixels with low methane concentrations are not filtered out. Second, uncertainties in the estimation for XCH_4 are based on the single sounding precision due to measurement noise. It is recommended to multiply the error by a factor 2, for more reliable results. Third, due to the need to use the low-accuracy sun glint measurements, observations over water are harder to interpret, combined with a lower coverage than measurements over land [6].

Table 2.11: CH₄ qa-value parameter definition.

qa-value	Condition	Remarks
1.0	Clear-sky	Highest quality data
0.8	Failed deconvoluted irradiance spectrum	Not pixel specific
0.4	Not clear-sky	
0.0	No convergence	Lowest quality data

2.5. Emission Quantification Methods for TROPOMI

Besides the method by Beirle et al. [1], two other mass-balance-based methods are used to quantify emissions from point sources: the Cross-Sectional Flux (CSF) and Integrated Mass Enhancement (IME) method, which are taken from Varon et al. [58]. However, the IME and CSF are only useful when a single orbit of TROPOMI data is considered. These methods are generally not applied for long term analyses, as opposed to the Beirle method. The CSF and IME methods are used to quantify emissions on an individual day. These quantification can be averaged over a longer period of time to obtain a long-term analysis.

2.5.1. The Beirle Method

The Beirle method has the ability to extract top-down emission maps at high spatial resolution. It combines TROPOMI NO₂ tropospheric columns V with wind fields \mathbf{w} . The Beirle model uses the temporal mean of horizontal fluxes to locate point sources. High gradients of fluxes can be found around these point sources, which improves the identification and localization performance.

The flux \mathbf{F} is calculated according to

$$\mathbf{F} = LV\mathbf{w} \quad (2.10)$$

where L is the ratio of NO_x/NO_2 . This ratio is taken as $L = 1.32 \pm 0.26$. According to the continuity equation, the divergence D of the flux \mathbf{F} can be used to obtain values for the sources E and sinks S , according to

$$D = \nabla(LV\mathbf{w}) = E - S \quad (2.11)$$

The sink term is dependent on a first-order time constant τ_c , which is taken as $\tau_c = 4 \pm 1.3$ hours. Since the divergence is a linear operator, long-term averaged emissions can be determined from the averaged sink term plus the divergence of the flux. The NO_x emissions can now be calculated as

$$E = S + D = LV/\tau_c + \nabla(LV\mathbf{w}) \quad (2.12)$$

The Beirle method uses offline TROPOMI Level 2 data products. Measurements with $qa < 0.75$ are not considered. TROPOMI NO_2 data are gridded on a uniform latitude-longitude scale with spacing 0.027° . To remove the upper tropospheric background and biases of the stratospheric estimate, the 5th percentile of the TROPOMI data is subtracted from the operational tropospheric column. The result is the lower tropospheric column V . Wind data is linearly interpolated between the measured location and the grid location. For the calculation of the divergence, numerical derivatives are calculated as the fourth-order central-finite difference, as shown in [Equation 2.13](#).

$$y' = \frac{y_{-2} - 8y_{-1} + 8y_{+1} - y_{+2}}{12h} \quad (2.13)$$

with h being the spacing in x , also defined as the spatial resolution. In this equation, y_i is the value at a certain gridpoint, where the subscript i denotes the location of the gridpoint with respect to the gridpoint of which the divergence is calculated.

Iterative Peak Fitting

To localize point sources and separate them from the background, iterative peak-fitting is used. This algorithm has three steps, as shown below. These steps are repeated until no more possible sources are found.

1. For a given map of E , the location with the maximum emission value E_{\max} is determined. The geo-location is saved as $(\text{lat}_{\max}, \text{lon}_{\max})$. This location is named P_{\max} .
2. Around P_{\max} , a 2D Gaussian is fitted.
3. The fitted peak is subtracted from the emission map E .

2.5.2. Cross-sectional Flux Method

The CSF method computes the flux through a cross section orthogonal to the plume axis to estimate the source rate. It makes use of the mass balance, which states that the source rate Q must be equal to the product of the wind speed U and the column plane transect along the axis perpendicular to the wind vector, according to

$$Q = \int_{-\infty}^{+\infty} U(x, y) \Delta\Omega(x, y) dy \quad (2.14)$$

where the integral is commonly approximated as a discrete summation of the product of the wind speed U and the vertical column enhancement $\Delta\Omega(x, y)$ over the plume width. An advantage of the

CSF method is that the method considers the full downwind extent of the plume. A disadvantage is related to the wind characterisation, which is the need to describe the vertical average of the wind velocity over the plume extent. This is problematic, since there is generally no information on the horizontal variability of the wind over the scale of the plume. However, this can be avoided by using a uniform effective wind speed U_{eff} applied to the cross-plume integral C , as shown in Equation 2.15. The integral C is independent of the downwind direction x when U_{eff} is taken as uniform with distance x downwind of the source. Now, C can be calculated for different values of x after which the values are averaged to improve the accuracy of the method [58].

$$Q = CU_{eff}, \text{ with } C = \int_{-\infty}^{+\infty} \Delta\Omega(x, y)dy \quad (2.15)$$

2.5.3. Integrated Mass Enhancement Method

The IME method relates the total plume mass detected downwind of the source to the source rate. The column plume IME is taken as

$$\text{IME} = \sum_{j=1}^N \Delta\Omega_j A_j \quad (2.16)$$

where N is the number of pixels of area A . The source rate calculation is based on the relationship between the IME and the residence time in the plume τ_p , which is a first-order time constant. The time constant can in turn be expressed in terms of the effective wind speed and the plume size L_p , according to Equation 2.17. Since the dissipation of the plume occurs through turbulent diffusion, U_{eff} and L_p must be considered as operational parameters, rather than simple physical constants. Since the detectable plume size L_p depends on the source rate and the wind speed, Equation 2.17 becomes non-linear [58].

$$Q = \frac{1}{\tau} \text{IME} = \frac{U_{eff}}{L_p} \text{IME} = \frac{U_{eff}}{L_p} \sum_{j=1}^N \Delta\Omega_j A_j \quad (2.17)$$

2.5.4. Supporting Data and Databases

To properly analyse and quantify local emissions, supporting data is used. The project uses wind data (GEOS-FP and ERA5), bottom-up emission inventories (EDGAR, REAS, TNO and DACCIWA) and the previously described cloud-data from Suomi NPP. The wind data is used in various emission localization and quantification methods, both for CTM-dependent and CTM-independent methods. In the method by Beirle et al. [1], it is used for flux calculations. In other methods, it is mostly used to analyse the source of plumes by analysing the wind direction or to calculate the source rate, as is the case for the IME and CSF methods. The emission databases are used to compare the results of the to be constructed model to known emission sources. Also, the model allows for evaluating the known emission quantities by comparing the model results with the values from the previously mentioned inventories. Furthermore, the project makes use of supporting data that is supplied with the TROPOMI Level 2 data. Possible data products that can be used are the surface pressure, surface albedo and the aerosol optical thickness.

ERA5

ECMWF ReAnalysis v5 (ERA5) is a reanalysis wind dataset that contains hourly wind data with a spatial resolution of $0.28^\circ \times 0.28^\circ$ ($\approx 31 \times 31 \text{ km}^2$). For the higher resolution grids, the same linear interpolation method as described in the Goddard Earth Observing System - Forward Processing (GEOS-FP) section can be used. The wind vectors are available at multiple altitudes, but this research uses the 10 m and 100 m altitude wind vectors. According to Olauson [59], ERA5 has a lower mean absolute error when compared to other similar wind datasets [60] Reanalysis datasets are

based on measurements, and some measurement and observation data points are within the same gridcell, with the majority of the gridcells not including any measurement station. To obtain global coverage, a data assimilation process is used to combine observations [61].

GEOS-FP Wind Data

GEOS-FP are wind data files supplied by NASA's Global Modeling Assimilation Office (GMAO). The wind direction and speed is provided at a $0.25^\circ \times 0.3125^\circ$ ($\approx 25 \times 25 \text{ km}^2$) grid. This resolution is reasonably high when considering a global scale. However, it is approximately a factor 10 larger than the standard resolution of 0.027° that is used in the method by Beirle et al. [1]. To estimate the relative wind direction and speed at the high resolution gridpoints, linear interpolation of the $0.25^\circ \times 0.3125^\circ$ can be used. The GEOS-FP wind data is available for every hour. The lowest grid point altitude is 60 m above the surface, but the wind speed at 10 m can be obtained by calculation from the 60 m wind speed, if necessary. On average, the GEOS-FP wind data has a standard deviation of 1.6 m s^{-1} at 10 m altitude, independent of wind speed [5]. In general, the surface wind speed is underestimated for low latitude (sub)tropical regions. For higher latitudes, the wind speed is mostly overestimated [62].

EDGAR

Emissions Database for Global Atmospheric Research (EDGAR) is a global bottom-up emission inventory with a $0.1^\circ \times 0.1^\circ$ grid resolution. EDGAR contains anthropogenic emission data for NO_x , CO and CH_4 for all 20 emission sectors (infrastructure, agriculture, etc.). EDGAR is generally used as the standard emission database in air quality modelling [63]. This research uses EDGAR data from the year 2017.

REAS

Regional Emission inventory in ASia (REAS) is a gridded bottom-up inventory of East, Southeast and South Asia. It includes emissions from fuel combustion in power plants, industry, transport and domestic sectors. It has a spatial resolution of $0.25^\circ \times 0.25^\circ$ and a monthly temporal resolution. The inventory includes estimates for industry including both combustion and non-combustion sources, road transport, other transport (non-aviation) and residential emissions. For this research, the average emissions from 2015 are used. The inventory includes both NO_x and CO [64].

TNO

To support policy related studies on air quality and climate modelling, TNO developed a spatially resolved non-gridded emission inventory at high resolution ($7 \times 7 \text{ km}$) on a European scale. The inventory is largely based on the official country emissions, since European countries report these to the EU and the UN. However, since the quality of these official emissions varies per country, the quantifications are studied in detail and appropriately adjusted. The inventory contains, among other, emissions for NO_x and CO [65] and is plant-specific, meaning only emissions from industry are included. This research uses data from the year 2017.

DACCIWA

The Dynamics-Aerosol-Chemistry-Cloud Interactions in West Africa (DACCIWA) inventory contains fossil fuel emissions for West-Africa. The inventory uses spatial proxies (geographical locations of plants and road networks) to convert national emissions into gridded inventories at a spatial resolution of $0.1^\circ \times 0.1^\circ$ for all emission sectors. It has a yearly temporal resolution and includes emissions for NO_x and CO. The inventory contains emissions up to 2015 [66]. Therefore, the year 2015 will be used in this research.

3

Research Definition

As stated in [chapter 1](#), CTMs are used for the most accurate estimation of emissions. However, although accurate, these models have the downside of having a high computational cost. This relates to both actual cost as well as processing time and the occupation of the computational structure. Therefore, this research investigates the use of models that do not rely on the use of CTMs. One of these models is the model by Beirle et al. [1]. Beirle developed a method for nitrogen dioxide that allows localization and quantification of point sources, without the use of a CTM.

In addition, it has the ability to not only estimate the emissions that are related to those specific point sources, but to calculate a full grid of emissions. This is a feature that separates the Beirle method from other CTM-independent methods. This thesis project aims to design an emission quantification and localization model for carbon monoxide and methane by means of a CTM-independent analysis method, which will be designed, optimized, validated and implemented for CO. The CTM-independent model for methane will only be developed, as the optimization and validation is out of the scope of this research. Both models will be based on the NO₂ model by Beirle et al. [1].

3.1. Relevance of the Project

The CTM-independent emission localization and quantification model for CO and methane can be used for many different types of research. It will be able to quantify emissions, and therefore can be used in case of a long-term blowout of a large super emitter. Furthermore, it is able to identify individual point sources, and therefore can be used in areas where there are multiple emission sources nearby. In case the model performs well and can localize and quantify emission sources adequately, the model can be widely used in research. The model will be specifically useful in the case where many sources are close-by. An example of this is shown in the Riyadh location elaborated in [section 1.2](#). This is made abundantly clear in [Figure 1.2](#), where multiple nearby sources can be identified in the emission map. Also, the model can be run for different spatial resolutions and data sets easily, which is more complicated when using a chemical transport model analysis. This allows for analysis using a wide variety of time periods, spatial resolutions and grid sizes. The main advantages of the model are the possibility to identify nearby emission sources and to obtain a full grid of emissions at a high resolution, without the need to run a CTM.

3.2. Research Question(s)

The main research question of this thesis is:

"How can the CTM-independent emission quantification and localization method for NO₂ by Beirle et al. [1] be adapted for carbon monoxide so it can be used for identification of closely-spaced emission sources and how can this model be optimized and utilized?"

The main research question will be answered by first focusing on the lower-level research questions. The answers of these sub-questions will together provide the answer to the main research question.

- How does the divergence model need to be tuned to accommodate the different characteristics of these gases?
 - How do the different lifetimes and mixing ratios of the gases effect the design of the divergence model?
 - How does the TROPOMI coverage of the three gases influence the performance of the divergence model?
 - What is the effect of different spatial resolutions on the output of the model, and what is the optimal resolution?
 - How can pseudo-observations be used to test the emission quantification estimates?
- Compared to other emission quantification methods, what errors are associated with the resulting emission fields and how sensitive are the results to the input data used and assumptions made?
 - What are the error sources for the divergence model?
 - What errors are related to the three gases and how does this differ for CO, NO₂ and CH₄?
 - How is the total uncertainty of the emissions estimated?
 - What is the effect of uncertainties in the meteo-data?
 - How well does the divergence model perform for different numerical differentiation methods?
- To what extent can these methods be used to separate the signals from large point sources from background emissions and quantify their emissions?
 - How well does the divergence model perform in identifying point sources when using a peak-fitting algorithm?
 - How well can nearby point sources be identified as individual sources, and up to which resolution is this possible?

4

Methodology

This chapter describes the methodology of the divergence method for carbon monoxide. It starts with a description of the data preparation, including the filtering and sub-pixel oversampling. After, it describes the different filters and corrections that are necessary to improve the model performance. It shows the Gaussian peak-fitting algorithm and the emission quantification based on this algorithm. Lastly, the adaptations for the methane model are briefly described. The chapter ends with the pseudo data that is used to test the performance of the method. A schematic overview of the divergence method is included in [Figure A.1](#) in [Appendix A](#).

4.1. Data Preparation

Before the TROPOMI data can be used as input for the divergence method, the data needs to be prepared. The TROPOMI data is filtered on domain (lat, lon) and gridded by a sub-pixel oversampling routine.

4.1.1. The TROPOMI Data Product

To process the orbit files into usable CO data files, first the orbit files are imported. All orbits that are within the specified time interval are selected. This means that the orbit files have global coverage within the time interval. After the data selection, the orbit files are processed. First, the necessary variables ([Table 4.1](#)) are extracted from the orbit files.

Table 4.1: The extracted variables from the TROPOMI Level 2 orbit files.

Variable	Unit
Time	[s]
Latitude	[degree]
Longitude	[degree]
XCO	[mol/m ⁻²]
Height scattering layer	[m]
Aerosol optical thickness	[-]
Solar zenith angle	[degree]
Processing quality flags	[-]
Latitude corners	[degree]
Longitude corners	[degree]
XCO precision	[mol/m ⁻²]
Surface pressure	[Pa]

Using these variables, the qa-value and the weekday are calculated. The calculation of these values is done in a previous stage and is therefore not part of this research. The qa-value is obtained by using the Aerosol Optical Thickness (AOT) and the Height Scattering Layer (HSL), as shown in [Table 4.2](#). The value qa=0.0 is used to filter out unusable data. The value qa=0.0 is obtained in two different ways. First, the TROPOMI data includes processing quality flags. If a pixel has a processing quality flag value which is unequal to zero, it obtains a qa-value of zero. Similarly, qa-values of pixels with a solar zenith angle larger than 80 degrees are set to zero as well.

Table 4.2: qa-values for corresponding AOT and HSL values.

qa-value	Limits	Description
1.0	AOT <0.5 HSL <500	Clear sky
0.7	AOT >0.5 HSL <5000	Medium thick clouds
0.4	All other values	High clouds Low thick clouds Medium thin clouds

After the data is filtered and corrected, the pixels are saved in a Comma-Separated Value (CSV) file. These files are processed by month and contain all data variables that are previously described. The processed global monthly CO CSV files are around 25 GigaByte (GB) in size, which makes importing these files in Python time consuming. Therefore, the monthly global CO files are re-selected into monthly local CO files. A list of coordinates (latitude, longitude) of to be analysed locations is used to select the corresponding domains. Around every location, a domain of $5^\circ \times 5^\circ$ is selected from the global data. These monthly local CO files vary from 10 MegaByte (MB) to 1 GB in size.

4.1.2. Data Selection

The locations in the location list are identified by a number. Two examples are given in [Table 4.3](#). When analysing a domain with a single point source, for instance a power plant at a remote location or a city that is assumed to be a point source, the latitude and longitude are taken as the same coordinates as the point source. In case a domain has multiple sources or is expected to have multiple sources, the grid size, resolution and latitude and longitude are chosen such that all sources are within the domain.

Table 4.3: Two examples to the input file.

Num	Latitude	Longitude	Name	Description
0	24.66	46.86	Riyadh	City center of Riyadh, steel and power plants in the vicinity
1	28.51	84.28	India	Jharkhand region in India, multiple large steel plants

Before a location can be analysed, certain parameters need to be set, as shown in [Table 4.4](#). The model grid resolution is equal to the FORTRAN resolution ([subsection 4.1.3](#)) and the FORTRAN resolution will be adjusted automatically when the model grid resolution is changed. The grid size is identical in the latitudinal and longitudinal direction, which results in a square grid. When a grid size of 50 gridpoints is chosen with a grid resolution of 0.05° , the full domain is a 101×101 grid with

an area of $5^\circ \times 5^\circ$. The centerpoint is independent of the grid size. The full length of the grid in the latitudinal and longitudinal direction is:

$$\text{Grid}_{lat} = \text{Grid}_{lon} = 2 \cdot \text{Grid}_{size} + 1 \quad (4.1)$$

Table 4.4: Parameters that need to be defined before a model run.

Parameter	Format	Description
Location number		Number of the location to select the corresponding latitude and longitude
Grid size		Grid length in number of gridpoints
Grid resolution		Degrees between adjacent gridpoints
Start date	DD-MM-YYYY	Initial date of the analysis
End date	DD-MM-YYYY	Final date of the analysis
qa-value		To filter out data below the selected quality value

Since the divergence method considers daily CO grids, the monthly local CO files are further reduced by filtering and data-selection. The pre-sampling process consists of four steps. (1) First, the data is filtered on latitude and longitude bounds and pixels with Not-a-Number (NaN) values in the XCO column (also called ‘bad pixels’) are filtered out. (2) The files are re-selected by date to obtain daily files. (3) Next, the model has an optional TROPOMI resolution filter, to filter out pixels with a large surface footprint which is common at the edge of the swath. (4) The resulting daily data files are re-selected based on orbit number. Some locations have double orbits for a single day, which results in overlapping data if the orbits are not separated. Overlapping data decreases the accuracy of the model since the data does not match the wind input for both orbits. Therefore, the initial daily orbit is saved by using an identifier: Single Orbit (SO) in the file name. The second orbit is identified by Double Orbit (DO). Both orbits are then used as input for the divergence model. The resulting local daily filtered CO files are significantly smaller than the initial files, with the daily files generally being less than 1 MB in size.

4.1.3. Pixel Oversampling

The FORTRAN sub-pixel oversampling routine is used to convert the TROPOMI data into a suitable grid that can be used for the divergence method. The routine needs two inputs: a TROPOMI data domain and an oversampling resolution. The centerpoint of the FORTRAN grid is taken as the approximate center of the domain, and the number of gridpoints depends on the width of the domain and the resolution. The FORTRAN oversampling method is shown in [Figure 4.1](#).

The figure shows a schematic representation of a TROPOMI satellite pixel p . The pixel p has an area of $S(p)$ and a value (XCO or surface pressure) of $\Omega(p)$. The black (dashed) boxes i represent the gridcells, which represent the borders around the gridpoints. The overlap between the TROPOMI pixel p and any gridcell i is denoted as $A(p, i)$. For any gridcell i , it has $N(i)$ overlapping satellite pixels. The $\bar{\Omega}(p)$ is the weighted average of the TROPOMI value within the gridcell of the pixel p . The average is calculated according to:

$$\bar{\Omega}(i) = \frac{\sum_{p=1}^{N(i)} \frac{A(p,i)}{S(p)} \Omega(p)}{\sum_{p=1}^{N(i)} \frac{A(p,i)}{S(p)}} \quad (4.2)$$

In case the gridcell has multiple overlapping TROPOMI pixels where one of the pixels has a NaN value, the routine does not use the NaN pixel for the calculation of the weighted average $\bar{\Omega}(p)$. The routine needs at least one TROPOMI pixel with a real value overlapping the gridcell to successfully

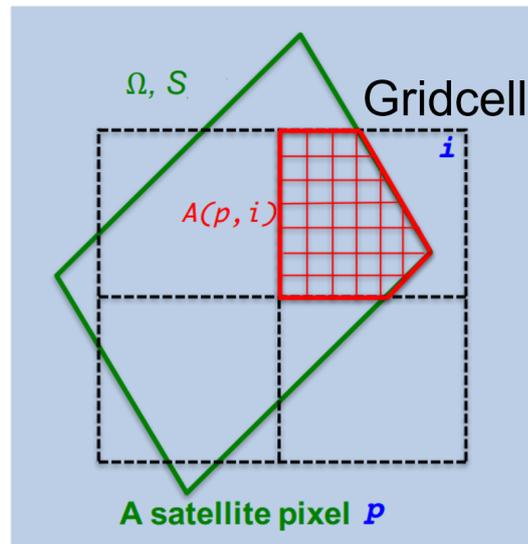


Figure 4.1: The FORTRAN sub-pixel oversampling method, showing four gridcells being overlapped by a TROPOMI pixel [67].

compute a weighted average. The FORTRAN sub-pixel oversampling routine has three main advantages. (1) It is computationally fast. The routine can sample a full year of data for a single parameter (XCO or surface pressure) in less than half a minute. The full routine for both variables then takes approximately one minute. (2) The routine fully uses and appropriately weights the information from all individual TROPOMI observations with a wide range of pixel sizes and parameter values. This makes the routine a robust option for oversampling. (3) The routine automatically grids the data and weights the values at the gridpoint. Therefore, there is no need to re-grid the data in a later stage. Also, since the data is weighted to calculate the value in each gridcell, there is less error associated with the gridding of the data.

The oversampling appropriately weights the information of each satellite pixel to calculate the value of the gridcell. However, when a gridcell is fully located within the boundaries of a single TROPOMI pixel, the value of the gridcell is equal to the value of the TROPOMI pixel. A special case arises when the domain contains missing pixels. In case a gridcell is overlapped by at least one TROPOMI pixel with a real value, it will obtain a value by the oversampling routine. A gridcell can only partly be overlapped by a TROPOMI pixel, while the residual of the gridcell is not overlapped by any pixel. This is shown schematically in Figure 4.2. In this figure, two TROPOMI pixels are shown with a low value (blue box) and high value (red box). The black boxes represent the gridcells in the domain. The oversampling routine weights the information each pixel to assign a value to each gridcell, as is shown by the gridcell colors. It can be seen that the bottom gridcells are fully weighted by the blue pixel, while the top gridcells are fully weighted by the red pixel. The top-left box is not overlapped by any pixel, and will therefore not be assigned any value. The area covered by the oversampled gridcells is now approximately twice as large as the area of the two TROPOMI pixels. Therefore, the oversampling routine increases the domain data coverage in case of missing pixels. In general, the divergence method uses grid resolutions similar to the TROPOMI pixel resolution.

The improvement of data coverage in the domain by the oversampling routine is dependent on the model resolution. This is shown in Figure 4.3, where the TROPOMI data for Lahore, Pakistan is can be seen on the left. In this sub-figure, rivers can be clearly identified by the string of missing pixels. Furthermore, the missing pixels in the top-right corner can be related to the Himalayas. The middle

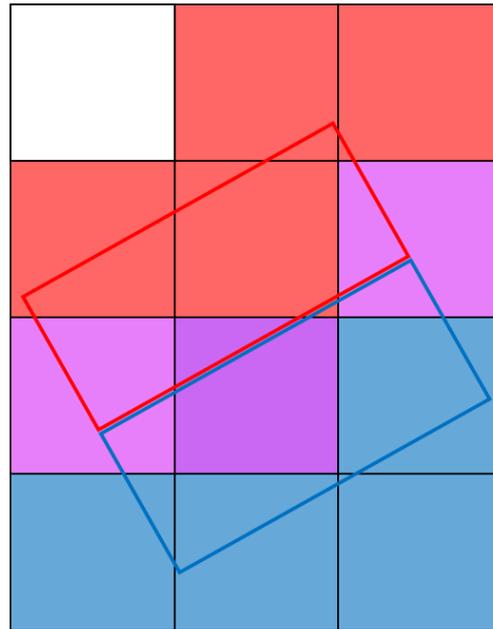


Figure 4.2: Two TROPOMI pixels, one with a low value (blue) and one with a high value (red) and the weighted average that is allocated to each gridcell. The white gridcell is not overlapped, and therefore results in a zero value.

and the right plot show the model grid coverage. A black gridcell indicates that the oversampling routine found a TROPOMI value in the gridcell. White gridcells indicate missing values. When using the oversampling routine on a low resolution of 0.1 degrees (middle), the missing pixels are mostly covered. When the model resolution is increased up to 0.02 degrees (right), the three rivers and the Himalayas can be clearly identified again. Therefore, when analysing low-coverage locations, running on a low resolution might be preferable. However, this negatively affects the accuracy of the data, since the TROPOMI data is smeared out over a larger area than initially observed.

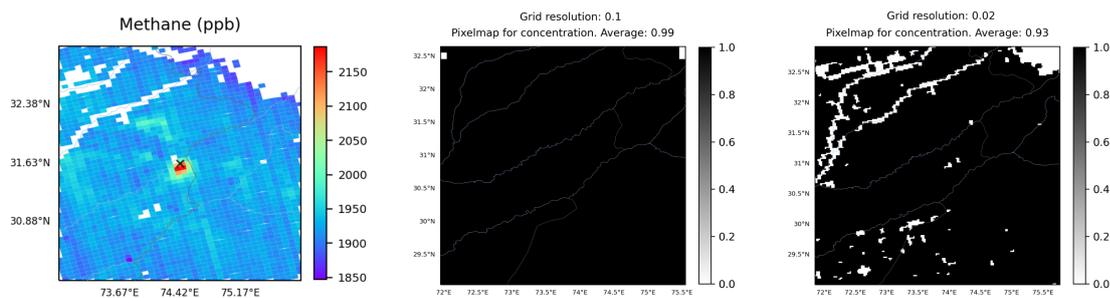


Figure 4.3: A methane plot (left) showing missing pixels over rivers. The oversampling routine improves coverage for low resolutions (middle) and partly improves coverage for high resolutions (right).

The FORTRAN files now consist of oversampled gridded CO data for a specified domain for a certain time interval. The domain of the FORTRAN gridded data is based on the grid that was created by the model, based on the latitude, longitude, grid size and resolution. However, the FORTRAN grid and the model grid do not necessarily overlap at each gridpoint. The resolution of both grids is equal, but the center-point often differs slightly. Therefore, if the two center-points do not match, the divergence model center-point is reset to match the center of the FORTRAN grid. The model grid is reset, since the FORTRAN grid contains measurements at specific latitude and longitude co-

ordinates, and therefore the data measurements will decrease in accuracy if shifted. Shifting the model grid, however, has no negative effects.

4.2. Model Functions

The daily gridded local CO files (subsection 4.1.3) are used as input for the divergence method. The CO and surface pressure values are directly assigned to the corresponding gridpoint in the model grid. Since the FORTRAN grid is slightly larger than the model grid due to the 0.5° buffer, not all the data is used. The FORTRAN CO data has the unit ppb. However, for emission quantification, g m^{-2} is a more convenient unit. This unit does introduce the issue of varying surface pressure, due to its dependence on the pressure p . The $CO_{\text{g/m}^2}$ can be calculated from CO_{ppb} using:

$$CO_{\text{g/m}^2} = CO_{\text{ppb}} \cdot \frac{M_{CO}}{M_{air}} \cdot \frac{p}{g} \cdot 10^{-6} \quad (4.3)$$

where M_{CO} is the molar mass of carbon dioxide (=28.01 kg/mol), M_{air} is the molar mass of air (=28.964 kg/mol), p is the surface pressure extracted from the FORTRAN grid (in Pa) and g is the gravitational acceleration constant (=9.80665 m/s^2).

4.2.1. Data Corrections and Filters

To improve the accuracy of the output and the performance of the model, data filters and corrections are used. The model uses one filter and three corrections:

- Land mask filter
- Altitude correction
- Background correction
- Interpolation of isolated missing values

Land Mask Filter

Similar to the TROPOMI NO_x data product, the TROPOMI CO data product includes measurements over water. However, due to the low surface albedo of water bodies, only scenes with cloud coverage can be used, since the top of the clouds have a higher albedo. The downside to these cloud-coverage measurements is the shorter atmospheric column. The TROPOMI CO concentration measurements are column-averaged XCO values. Therefore, the height of the atmospheric column is one of the influencing factors of the measured amount of CO in the atmosphere. Since the column is shortened by not including the part of the atmospheric column below the cloud, the CO concentration measurement with cloud-coverage will be lower. Furthermore, the part of the atmosphere where most emissions take place is not observed by TROPOMI when only the column above the cloud is used, which again results in unrealistic low measurements.

Due to the varying altitude of the top of clouds, or aerosol Height Scattering Layer (HSL), the CO concentration measurements over water are hard to interpret. Furthermore, due to the large variation in measured CO values, the divergence method detects large erroneous emissions over water bodies. Another problem arises at coastal regions. The CO values on water-based gridpoints are lower, as compared to the land-based gridpoints. The divergence method detects this variation and the output shows large divergence values at the coast. Since the wind direction in coastal regions is mostly land inward, the divergence method identifies the CO increase as downwind, which leads to a positive divergence value at the coast. The gridpoints with a positive divergence are in turn wrongfully identified as emission sources.

To resolve this problem, a land-mask filter is used. This process filters out all gridpoints over water. Large bays and lakes are included in the land-mask. However, small lakes and rivers are not identified by the filter. For instance, the IJsselmeer region in the Netherlands is not identified as water body, while the Bay Area (San Francisco) is included in the land-mask. First, the land surface is imported from Cartopy's Natural Earth Feature. For the filter, a resolution of 10 m is used. Next, the land body is defined as one or multiple polygons, dependent on the shape of the land body, to further simplify the process. Now, the filter tests whether each gridpoint in the domain is inside the polygon(s), defined by the previous step. If not, the gridpoint is filtered out.

The land-mask filter decreases erroneous emission values in coastal regions, as well as over water bodies. The major downside to the filter is the excessive data loss in coastal regions. Since data above water is filtered out, the two gridpoints neighbouring the coastal grid edge are not usable for the divergence calculation. A result of this data loss is that coastal emissions sources cannot be analysed by the divergence method.

Altitude Correction

The influence of the height of the atmospheric column on the column-averaged CO concentration measurement results in another altitude-related issue. Similar to measurements above clouds, CO concentration measurements at increased altitude above sea level result in lower CO values. This problem is similarly caused by the shortened atmospheric column. As the surface height above sea level increases, the measured CO value decreases. For simplicity purposes, this relation is assumed to be linear. Domains that have a large variation in surface elevation experience erroneous values in the divergence calculation. The altitude-CO relation can be seen clearly in Figure 4.4. In this plot, the Jharkhand steel plants in the mid-East of India can be seen, as well as the larger cities (New Delhi in the north, Mumbai and Hyderabad in the south). Furthermore, in the north of the grid, Nepal (and the Himalaya) can be seen. Due to high elevation of Nepal, on average 3,265 meter, the CO concentration measurement over this part of the domain is significantly lower as compared to the India lowlands. Next, the river-regions in India can be seen to have an increased CO enhancement. This is partly due to the higher pollution levels due to the high population density in these areas, as well as the lower surface elevation of the rivers, as compared to their neighbouring areas.

Surface pressure - CO concentration relation

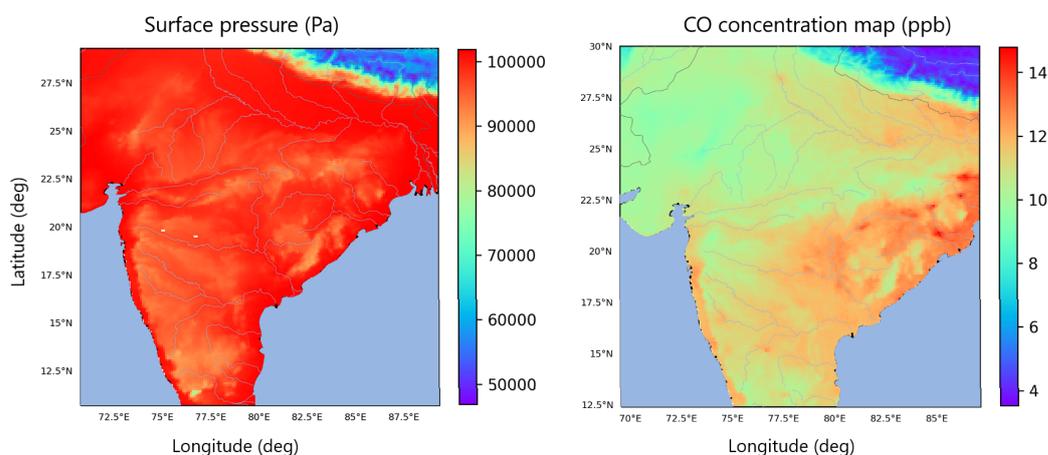


Figure 4.4: Average CO concentration map (g m^{-2}) for two months in early 2019, showing India and Nepal. Extremely low CO concentration values can be seen for low surface pressure (high altitude) around the Himalayas.

Since the altitude-CO correlation is assumed to be linear, the altitude correction formula is:

$$CO_c = CO_m + h \cdot \beta \quad (4.4)$$

where CO_c is the corrected CO value, CO_m is the TROPOMI measured CO value, h is the altitude above sea level and β is the altitude correction factor. The altitude can be calculated from the surface pressure, which can be extracted from the TROPOMI data product, according to:

$$h = \ln\left(\frac{p}{p_{atmos}}\right) \cdot \frac{-1}{0.12} \cdot 10^{-3} \quad (4.5)$$

where p_{atmos} is the atmospheric pressure at sea level, taken as 101,325 Pa. The altitude approximation formula in Equation 4.5 is valid up to 11 kilometers, which makes it sufficient to use for surface elevation calculations. To correct the data for this altitude correlation, the altitude correction factor (β) needs to be determined. This factor is calculated for every day. Linear regression is used to find the slope of regression line, as shown in Figure 4.5. The slope is then taken as the altitude correction factor.

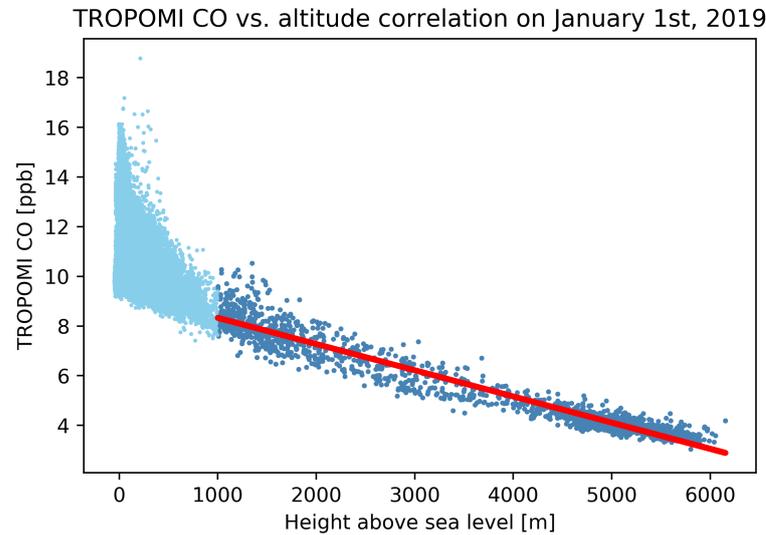


Figure 4.5: The altitude-CO relation for India on 2019-01-01. The dots denote the CO concentration measurement at a certain altitude, the red line is the linear regression line. The TROPOMI CO value denotes the oversampled CO values with a fifth percentile background correction. CO concentration measurements below 1,000 m (light blue) are not used for the calculation of the altitude correction factor.

In Figure 4.5, it can be seen that below 500 meters of altitude, the CO concentration measurements do not behave linearly when compared to altitude. This is due to CO emissions at these altitudes, since many steel plants in India are located below 500 meters of altitude. Therefore, increased CO concentration measurements can be observed in this altitude region. As can be seen in Figure 4.5, these CO emissions results in a slight over-estimation of the actual slope. Therefore, it is necessary to filter out areas with CO emissions, to only look at the background CO level. This, however, is not a straightforward action. Even though the surface altitude at which steel plants are located are mostly known, the surface altitude at which the emissions are observed is not related to the surface altitude of the steel plants. Due to changing winds, the CO emissions can be observed at various surface elevations. Therefore, it is not sufficient to filter out altitudes at which emission sources are located. In the case of India, the CO is observed to be emitted below 500 meters of altitude. However, CO enhancement that deviate from the linear regression line are observed up to 1,000 meters of surface altitude. Therefore, all CO concentration values below 1,000 meters are filtered

out (light blue markers in Figure 4.5). The linear regression uses only the CO concentration values above 1,000 meters of altitude (dark blue markers). The corresponding slope is taken as the altitude correction factor.

The altitude correction factor is not constant throughout the year. Due to the a varying tropopause height, there is a seasonal change observed in the altitude correction factor. The height of the tropopause determines the relative weights of the stratosphere and troposphere for the XCO calculation. The surface elevation results in a reduction of the mass of the tropospheric sub-column. Therefore, an increase in surface elevation shifts the contribution of the tropospheric and stratospheric air mass to the total column. Since the CO abundance is lower in the stratosphere compared to the troposphere, the measured XCO decreases as altitude increases [68]. When analysing the daily altitude correction factors for a full year, this variation can be clearly seen (Figure 4.6).

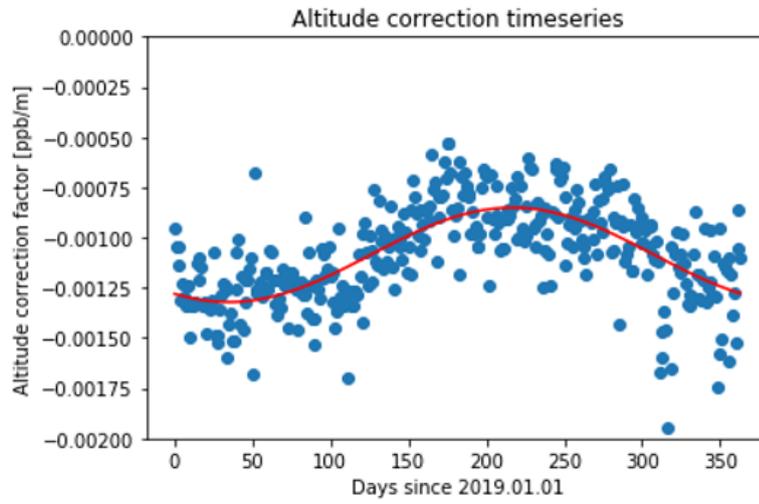


Figure 4.6: The seasonality of the altitude correction factor. The red line represent the fitted average.

To obtain a seasonal altitude correction factor, a sine is fitted to the data. To fit the curve, a Harmonic Oscillator hybrid fit (HOBIT) function is used. This function fits the data according to:

$$f(x) = y_0 + y_1 \cdot \sin(\omega \cdot x + \phi) \quad (4.6)$$

First, the algorithm splits the data into random train and test subsets. Next, the training data is used to find a fit for the data using a least squares approximation. The function will return the best-fit values for all parameters in Equation 4.6. To test the fit, the result of the data fit is used on the test data, after which the Root-Mean-Square Error (RMSE) is calculated:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (4.7)$$

where \hat{y}_i is the predicted value of observation i , y_i is the actual value of observation i and n is the number of observations in the data. The value of the sine function (Equation 4.6) is then used as the altitude correction factor for a specific day. An example of the altitude correction factor values can be found in Figure 4.6.

Background Correction

The divergence method uses the first-order derivative of the flux to calculate the emissions in a grid. Since the divergence uses the derivative of the flux, the method is able to identify changes in the flux field. The flux is dependent on two variables: the amount of CO in the atmosphere and

the wind speed (and direction). A change in the flux field can therefore be caused by a variation in both the CO and the wind field. However, to appropriately estimate emissions, only changes in the CO concentration should be considered. The problem that arises from the influence of the wind on the flux, and therefore the divergence, is illustrated in figure [Figure 4.7](#). This figure shows four subplots, from left to right: (1) an artificial input signal named 'Background', (2) the GEOS-FP 10 meter altitude wind speed in x -direction (longitudinal), (3) the corresponding wind speed in the y -direction (latitudinal), and (4) the corresponding divergence output.

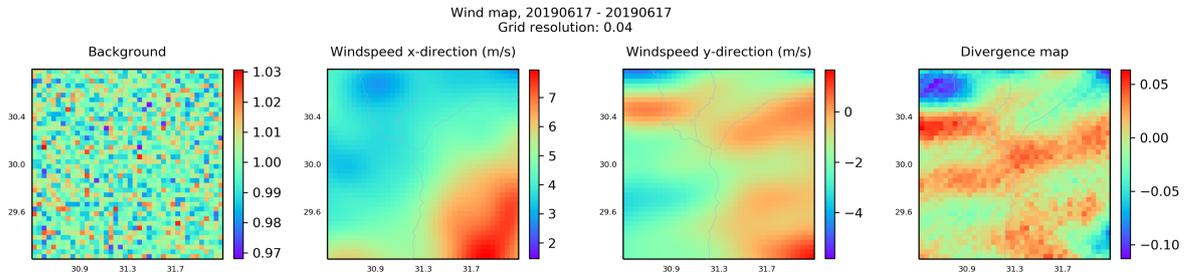


Figure 4.7: A simulated case without background correction. Enhancements in the divergence map are the result of the wind speed in combination with a high background.

The input signal is taken as random Gaussian noise, with $\mu = 1$ and $\sigma = 0.01$. Therefore, there is only a small deviation input values (min = 0.97, max = 1.03). This input signal is reasonably similar to the TROPOMI CO data that is used to calculate the divergence. Since the input is solely Gaussian random noise, the divergence output is expected to be similar to the input signal, and behave more or less random. However, the divergence map shows clear enhancements. Furthermore, the divergence output is strongly correlated to the wind speed. To counter this wind-induced problem, a background correction is used. To remove the background, the fifth percentile of the CO data is calculated. This value is subtracted from the CO data, after which all negative values are filtered out. By removing the background, only the CO enhancements are used for the flux calculation. This minimizes the influence of the wind on the flux. To illustrate the background correction, the previous example is used with a new input of random Gaussian noise with $\mu = 1$ and $\sigma = 0.3$ ([Figure 4.8](#)). By using a larger variance, the lowest values are observed to be around zero. This simulates the background corrected TROPOMI CO data.

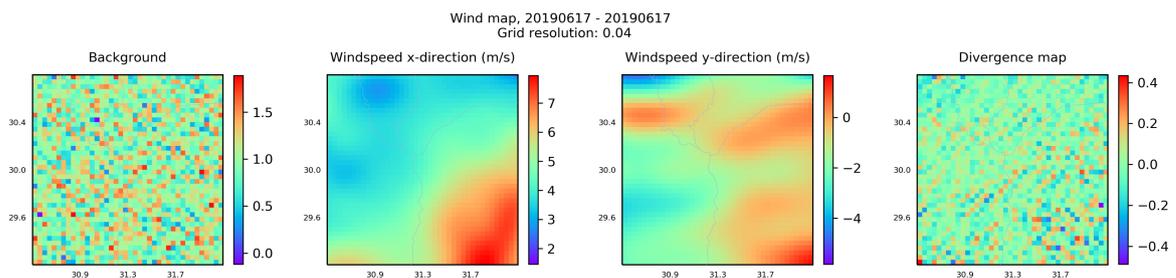


Figure 4.8: A simulation for the background correction. The erroneous divergence enhancements are mostly fixed.

When looking at the divergence map with this new input, no correlation with the wind fields can be observed, and the output shows mostly random behaviour. By using a fifth percentile background correction, the influence of the divergence of the wind field on the total divergence is minimized.

Minimising Data Gaps

One of the main causes of data loss in the divergence output is individual gridpoints with missing CO concentration values. One gridpoint without a concentration value will cause 8 neighbouring

gridpoints to not have a divergence value, since the derivative calculation is insufficient. This is shown in Figure 4.9. The left sub-figure shows the pixelmap (gridpoints that have coverage) for the CO concentration. The right figure shows the corresponding pixelmap for the divergence. Around the missing gridpoints in the concentration pixelmap, a plus-like structure can be seen in the divergence pixelmap. A small number of individual missing gridpoints in the concentration map result in extensive data loss for the divergence map. In this case, approximately 4% of the grid coverage for the divergence is lost due to the background correction, 2% due to multiple neighbouring missing gridpoints and 2% due to individual missing gridpoints.

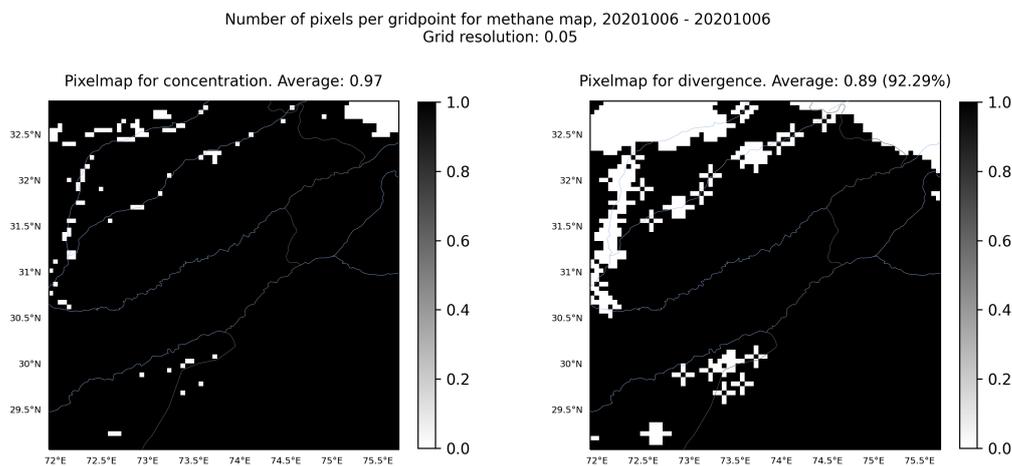


Figure 4.9: The coverage map, showing grid coverage (average of TROPOMI pixels per gridcell) for concentration (left) and coverage for divergence (right). The data loss by individual missing pixels ('+' form) can be seen in the divergence map.

To minimize this data loss, individual gridpoint with missing concentration values are linearly interpolated with their nearest neighbours. In case a gridpoint with a missing value is neighbouring another gridpoints with a missing value, the value is not interpolated. The diagonal neighbouring gridpoints are not included in the interpolation of a gridpoint, the method uses only the four non-diagonal neighbouring gridpoints.

4.3. Model Analysis

The divergence is calculated for each gridpoint separately for every single day. In some cases, locations have double orbits on one day. Due to overlap between swaths in adjacent orbits, one domain might be included in more than one orbit, as shown in Figure 4.10. This occurs mainly at high latitudes. The figure shows the revisit frequency for different locations on the Earth. Low latitude domains experience a double orbit only sporadically, while higher latitude can have up to seven orbits per day.

In case a domain has perfect coverage (no missing data) and no double orbits, the coverage map will show 365 pixels per gridpoints for a full year of data. This number can increase when domains have multiple days with double orbits per year. To start the model analysis, all grid files for every single day are imported, including days with multiple orbits. For every gridpoint, the mean values of CO concentration and divergence are calculated for the full time interval. This process consists of four steps: (1) one orbit grid file is imported. (2) Each gridpoint with a value in the orbit grid file is added to the mean grid. (3) The model checks how many data entries (v) are used for the sum of the values for one gridpoint. (4) The mean is calculated by dividing the sum of the values by v . In

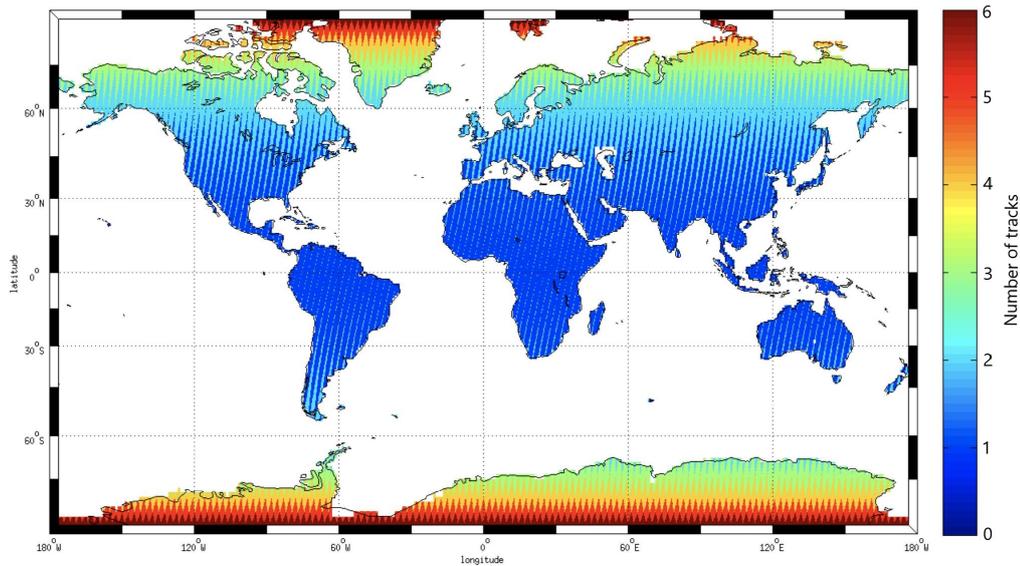


Figure 4.10: The daily geometric revisit frequency of TROPOMI.

general, every additional data entry leads to a value of ν according to $\nu = \nu + 1$.

One problem arises for low-coverage grids. Generally, the model output is unstable for gridpoints with a low number of pixels used to calculate the mean. In Figure 4.11, the value of the divergence and the number of pixels for each gridpoint are plotted. The divergence value becomes unstable for $\nu < 20$. Therefore, all gridpoints with $\nu < 20$ for the divergence are filtered out in the temporal mean.

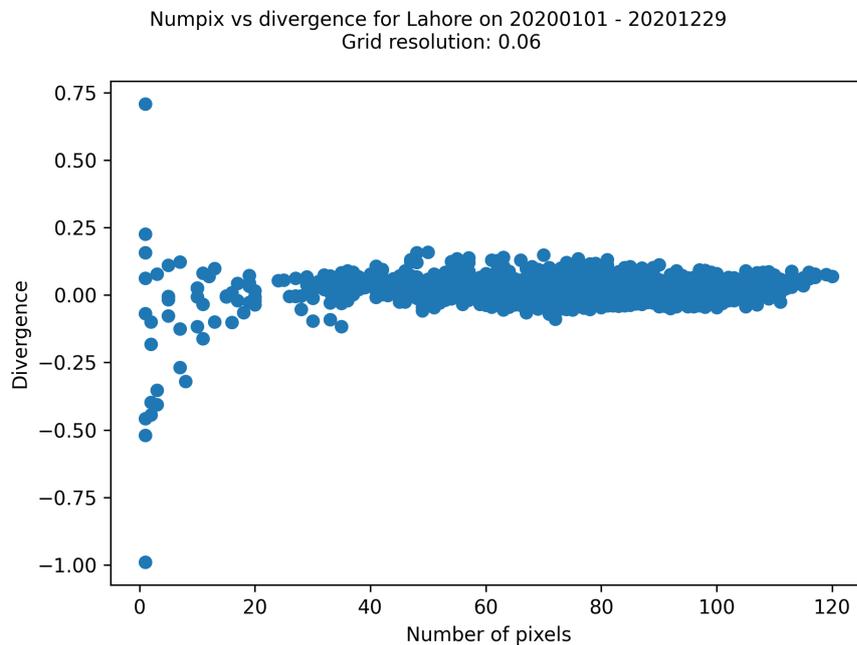


Figure 4.11: The model output showing gridpoints with their number of pixels and the corresponding divergence. The output becomes stable above 20 pixels.

The divergence method relies on good coverage to have a stable output. However, the TROPOMI coverage is mostly dependent on the geographical location of the grid. For instance, when a grid

containing only the Netherlands is used for the divergence analysis with a time interval of October 1st - March 31st, the coverage is expected to be low. This is due to the high cloud coverage in these autumn - winter - spring months. Most likely, a large part of the grid will not have good (>20 pixels) coverage (Figure 4.11). Therefore, this location in combination with the time interval is not feasible. However, this does not mean that all locations with a winter time interval are infeasible. Riyadh, for instance, has sufficient coverage year-round and therefore can be used in combination with a time interval containing only the winter months. One solution to the low-coverage locations is to simply use a larger time interval. Mostly, using a full year of data is sufficient.

4.3.1. Divergence Calculations

After the TROPOMI CO data is re-selected, oversampled, filtered and corrected, the divergence can be calculated. Wind data from GEOS-FP and ERA5 is used for the divergence calculations. Both wind fields have a resolution of 0.25°. However, to properly calculate the divergence, the wind vector at every gridpoint is needed. Therefore, both wind fields are spatially interpolated to find the wind vectors at intermediate grid points. The model uses a spline interpolation function for the wind interpolation. Spline interpolation is used since it has a reasonably fast computational time and favorable accuracy as compared to linear interpolation, which has a faster computational time but lower accuracy.

Now, the flux $\mathbf{F} = (F_x, F_y)$ is calculated by using the wind vector $\mathbf{w} = (u, v)$ and the CO concentration column value (Ψ):

$$\mathbf{F} = \Psi \cdot \mathbf{w} \quad (4.8)$$

The flux is calculated in the x -direction by using the u wind vector and in the y -direction by using the v wind vector. To calculate the divergence, first the spacing of the grid in the x - and y -direction is needed. To obtain these variables, the dimensions of the grid cells can be used. Since the length of 1° in meters in longitude varies with latitude, it is calculated for every gridpoint separately. To calculate the distance d in meters from the grid resolution in degrees, the haversine ($\text{hav}(\theta)$) function is used. The haversine determines the great-circle distance between two points on a sphere, as shown in:

$$d = 2R \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \left(1 - \sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) - \sin^2 \left(\frac{\varphi_2 + \varphi_1}{2} \right) \right) \cdot \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (4.9)$$

where φ is the latitudinal coordinate, λ is the longitudinal coordinate and R is the nominal radius of the Earth in meters ($R = 6,371,000$ [m]). To calculate the length of the gridcell in the x (longitudinal) direction, the $\Delta\lambda$ ($\lambda_2 - \lambda_1$) is taken as twice the grid resolution, while the $\Delta\varphi$ ($\varphi_2 - \varphi_1$) is taken as zero. To calculate the length of the gridcell in the y (latitudinal) direction, $\Delta\varphi$ is taken as twice the grid resolution, while $\Delta\lambda$ is taken as zero. The gradient of the flux, or the divergence (D), can be calculated as the first-order partial derivative of the flux in the x and y direction, according to:

$$D = \nabla \mathbf{F} = \frac{\partial}{\partial x} F_x + \frac{\partial}{\partial y} F_y \quad (4.10)$$

The spacing value (distance d) is then used to calculate the divergence in the x - and y -direction. The calculation uses a fourth-order central finite difference numerical differentiation method (Equation 4.11) to calculate the derivative of the flux.

$$D_0 = \frac{-F_{-2} + 8F_{-1} - 8F_{+1} + F_{+2}}{12h} \quad (4.11)$$

In this equation, the subscript of F denotes the gridpoint that is used for the calculation, with respect to the gridpoint for which the divergence is calculated, and F is the flux value at the gridpoint.

The divergence therefore uses two gridpoints west and two gridpoints east when calculating the divergence in the x -direction. For the calculation of the divergence in the y -direction, it uses two gridpoints north and two gridpoints south of the considered gridpoint (subscript 0). In [Figure 4.12](#), the grid is schematically shown. To calculate the divergence of the flux at a certain gridpoint (red), four neighbouring gridpoints are used (blue).

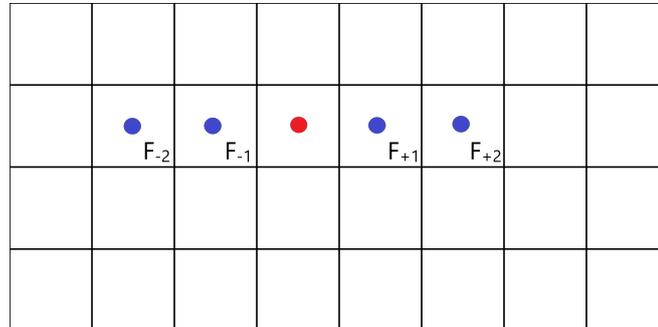


Figure 4.12: A cut-out from the model grid, showing the gridcells (black), the gridpoint used for the divergence calculation (red) and the gridpoints used for the numerical derivative (blue).

One disadvantage of the fourth-order central finite difference numerical derivative method is the extensive data loss. As stated before, the two rows and gridpoints closest to the grid edge can not be properly calculated, since there is not enough data to calculate the derivative. This problem can be overcome by increasing the grid size by 2 gridpoints to run the previously defined grid size. However, the grid edges are not the sole reason for data loss due to the numerical derivative. Data gaps exist over land-based water bodies such as rivers and lakes, as well as large water bodies such as seas and oceans. Therefore, coastal regions experience data loss similar to the grid points neighbouring the grid edge. In areas with lakes and rivers, the data loss is often limited to a single or multiple TROPOMI pixels. It then depends on the grid resolution as to what number of grid points with data loss this then translates to. An example is illustrated in [Figure 4.13](#). In this figure, the three gridcells with the red dot represent a missing TROPOMI pixel, and therefore have no assigned CO value. Since the numerical derivative method needs the neighbouring gridcells to have an assigned CO value to calculate the divergence, the missing gridcells result in extensive data loss around the missing gridcells. In this example, approximately 10% of the gridcells are have a non-existing CO value. This results in approximately 50% of the gridcells having non-existing divergence values (red gridcells). In the figure, only the green gridcells can be properly calculated.

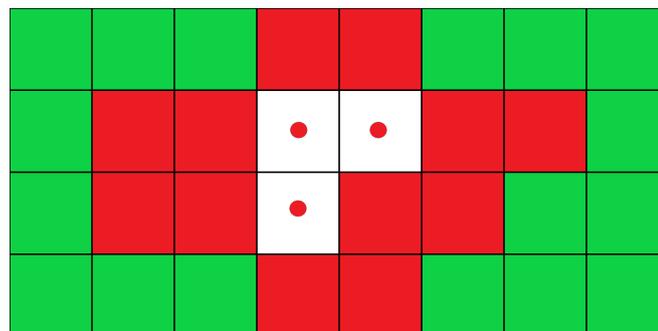


Figure 4.13: A cut-out from the model grid, showing the gridcells without a TROPOMI CO value (blank with red dot). All other gridcells do have a TROPOMI CO value. This data gap results in extensive data loss (red gridcells) for the divergence calculation. The green gridcells do have a divergence value.

The spacing h_x is equal to the width of gridcell and the spacing h_y is equal to the height of the gridcell, as previously calculated. As all variables are known, the divergence can be calculated. This

is again done in the x - (Div_x) and y -direction (Div_y) separately, and for every gridpoint. Since the calculation of the divergence needs two gridpoints with known values in all four cardinal directions, the points closer than three gridpoints from the grid edge can not be properly calculated. Therefore, these are not taken into consideration. After all gridpoints are calculated, the two outer rows and columns of gridpoints are filtered out. To obtain the total divergence at every gridpoint, the calculated divergence values Div_x and Div_y are added up.

4.3.2. Gaussian Peak-fitting

The mean divergence map which is calculated in the previous section is now used to calculate the emissions. This is done, similarly to the divergence calculation, for every gridpoint. Since the divergence (D) has the unit [$\text{g m}^{-2} \text{s}^{-1}$], the emissions E (Gg a^{-1}) per gridcell can be easily calculated by using:

$$E = D \cdot A \cdot t \cdot 10^{-9} \quad (4.12)$$

where A is the area of the gridcell (m^2) and t is the time interval (in years). The area is calculated by using the width and length of the gridcell, according to Equation 4.9 on page 37. The resulting emission map is used for the iterative Gaussian peak-fitting algorithm, which consists of three steps:

1. Identify the gridpoint with the highest value in the emission map. The location of this gridpoint is identified as (P_{\max})
2. Fit a 2D Gaussian on the gridpoint at P_{\max} .
3. Subtract the fitted signal from the emission map.

This is an iterative process. Before the process is carried out, the number of iterations for the Gaussian peak-fitting algorithm must be defined. In the emission map, there are two types of high values: sources and artefacts. Artefacts are locations that show high CO emission enhancements, without emission sources in the vicinity. Emission sources mostly have spatial Gaussian behavior in the emission map. For this reason, the algorithm can fit a Gaussian on top of the gridpoint with the highest value. However, artefacts generally do not have a Gaussian form in the spatial domain. Therefore, the algorithm generally can not fit a Gaussian on top of artefacts. If the algorithm is not able to obtain a proper fit, the corresponding gridpoint will be identified with a *skip*. This results in the value being skipped when the algorithm is looking for the highest value in the grid for the next iteration. This way, the algorithm will not stay in an infinite loop when identifying artefacts.

After the gridpoint with the highest value (non-Skip) is identified, a 2D Gaussian is fitted on the data around the gridpoint, according to

$$f(x, y) = H \cdot \exp \left(- \left(\frac{(x - x_0)^2}{2\sigma_x^2} + \frac{(y - y_0)^2}{2\sigma_y^2} \right) \right) \quad (4.13)$$

where H is the height of the Gaussian, x_0 and y_0 are the coordinates of the center and σ_x and σ_y are the widths of the Gaussian. To fit the Gaussian, first the data is normalized. The data is normalized to be able to enforce limits in the fit of the Gaussian, especially in terms of the height H . Then, the fitting process uses a non-linear least squares optimization process to find the best parameters for the Gaussian fit. The least squares method adjusts the parameters of the Gaussian fit to best match the model data.

The method selects the fit by minimizing the sum of squared residuals S_s for all n gridpoints in the grid, according to Equation 4.14.

$$S_s = \sum_{i=1}^n E_i - f(p_i)^2 \quad (4.14)$$

where E_i is the value of the emission data at gridpoint i , $f(p_i)$ denotes the value of the Gaussian function at gridpoint i , and p_i denotes the values of the parameters for gridpoint i to construct the Gaussian. The necessary parameters are (1) the height H (normalized to 1), (2) the width in x -direction, (3) the width in y -direction, (4) the x -coordinate of the center, and (5) the y -coordinate of the center. Since the data is normalized, the initial guess for the height of the Gaussian is taken as one. The initial guesses for the x - and y -coordinate of the center of the Gaussian are taken as the coordinates of the previously selected gridpoint. The initial guess parameters for the width are taken as zero.

To force the method to fit the Gaussian to the selected peak (P_{\max}), bounds are used for the least squares optimization. Firstly, the most strict bounds are used for the center of the Gaussian. To make sure no other peaks are fitted, the center of the Gaussian is forced to deviate no further than one gridpoint from the gridpoint P_{\max} . For the height of the Gaussian, the method is forced to stay within 5 percent of the peak height: $H = [0.95, 1.05]$ to make sure the algorithm fits the desired peak in the data. Lastly, since the width of the Gaussian varies per location and model resolution, no bounds are used for these parameters.

To obtain the best possible fit, the Gaussian peak-fitting algorithm uses a rotation for the fit. This sequence uses a loop that runs the Gaussian fitting for a full 360 degrees rotation, with 1° intervals. For every degree of rotation, the RMSE (Equation 4.7) of the fit is calculated. The degree of rotation with the lowest RMSE is selected for the final fit. To include a rotation in the fit method, the Gaussian function is expressed as:

$$f(x, y) = H \cdot \exp\left(-\left(a(x - x_0)^2 + 2b(x - x_0)(y - y_0) + c(y - y_0)^2\right)\right) \quad (4.15)$$

Now, the rotation parameters a , b and c are expressed as:

$$a = \frac{\cos^2 \theta}{2\sigma_X^2} + \frac{\sin^2 \theta}{2\sigma_Y^2} \quad (4.16)$$

$$b = -\frac{\sin 2\theta}{4\sigma_X^2} + \frac{\sin 2\theta}{4\sigma_Y^2} \quad (4.17)$$

$$c = \frac{\sin^2 \theta}{2\sigma_X^2} + \frac{\cos^2 \theta}{2\sigma_Y^2} \quad (4.18)$$

where θ is the clockwise rotation angle. After the best possible fit is determined, the fitted Gaussian is subtracted from the emission map, after which the next iteration will take place. This is repeated, until the pre-defined number of iterations is reached. All fitted Gaussian outputs are merged in a new emission map, which is used for further analysis.

4.3.3. Emission Quantification

After the model carried out the pre-defined number of iterations for the Gaussian peak-fitting algorithm, the model output is compared to bottom-up inventories. Bottom-up inventories estimate emissions using statistical analyses of activity data combined with country-specific emission factors, as opposed to top-down methods, which estimate emissions from observations [69].

For the CO model analysis, three different inventories are used:

- EDGAR
- REAS
- TNO
- DACCIWA

The EDGAR database consists of world-wide estimates of CO emissions, while REAS is limited to Asia and the Middle-East, TNO is limited to Europe and DACCIWA is limited to Africa. Even though the REAS and TNO database are limited in domain, the spatial accuracy of the inventories is higher compared to EDGAR. In Figure 4.14, an example of the emission output is shown. In the left subplot, the divergence model emissions are shown, including three known locations of steel plants in the Jharkhand (India) area, denoted by the markers. The right subplot shows the emissions from the REAS inventory. The estimates are shown in averaged yearly emissions per gridcell. For the REAS emissions, the area of the gridcells is $0.25^\circ \times 0.25^\circ$. For the model emissions, the area of the gridcells is $0.03^\circ \times 0.03^\circ$.

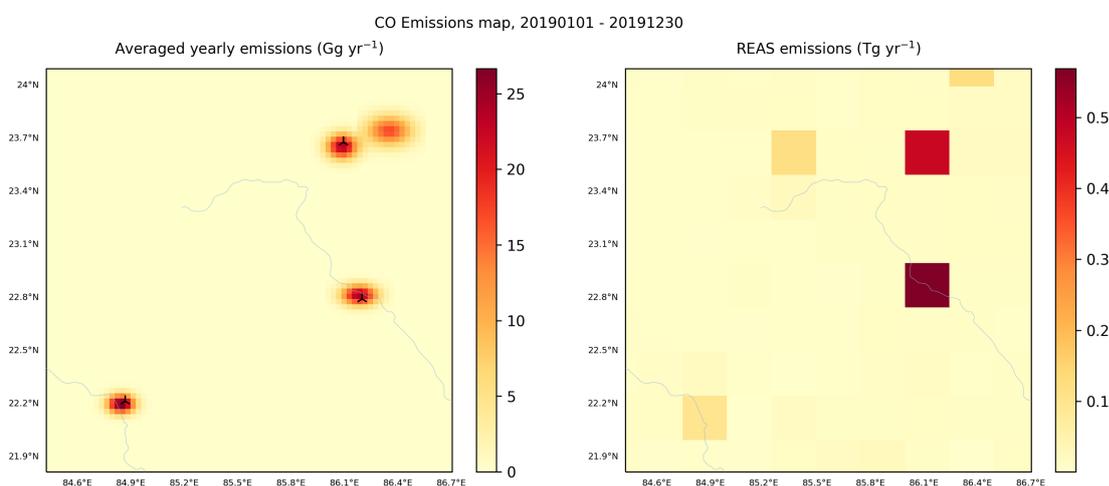


Figure 4.14: The emission map, showing the identified sources by the Gaussian peak-fitting algorithm (left), as well as the emission estimate from the REAS inventory for Jharkhand, India. All sources match known emission sources.

Artefact Identification

In Figure 4.14, four point sources have been identified by the Gaussian peak-fitting algorithm. However, there are only three known sources in the area. The resulting source can now be one of two things: either it is an unknown point source, or it is an artefact. This can be determined by looking at the CO concentration on individual days. If the identified source location shows plumes being emitted from that location, then it is most likely that the source is an emission source. If no plumes are observed, but the individual days do show enhancements which can be related to orography or waterbodies, it is most likely an artefact. To identify plumes, not only the CO concentration map needs to be taken into consideration. Plumes are known to deviate with the wind direction, always elongating in the downwind direction. Therefore, by looking at both the CO concentration map and the wind field, it can be determined whether an enhancement signal is a plume.

By looking at [Figure 4.15](#), it is very likely that the Jharkhand area has multiple sources. The markers indicate some of the larger sources, but some smaller sources have not yet been identified. It is clear that the previously described location is indeed an emission source.

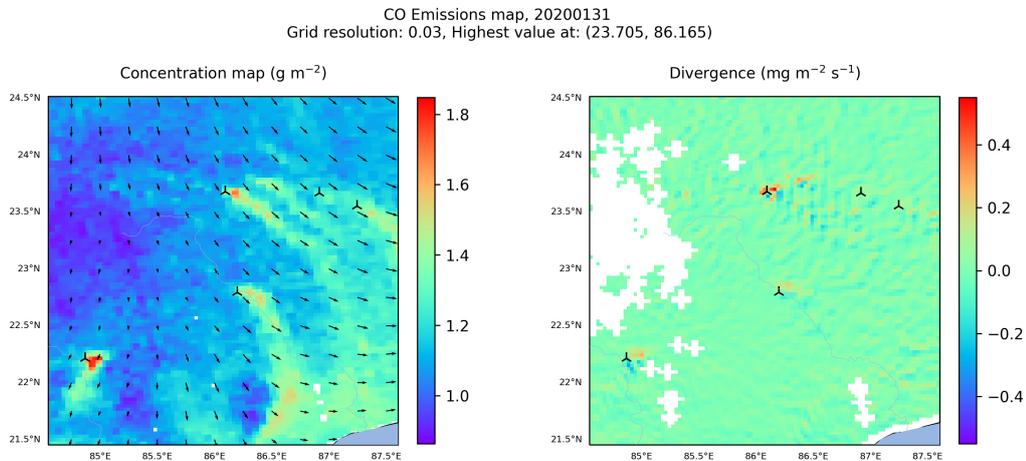


Figure 4.15: The CO concentration map for January 31st, 2019 with the GEOS-FP 10 m wind field (left), showing multiple plumes in the Jharkhand area. The wind speed and direction is shown by the black arrows. The resulting divergence output is shown on the right. Known emission sources are marked.

4.3.4. An Explanation For Divergence Output Form

The divergence method uses a 2D Gaussian (latitudinal, longitudinal) for the fit on the TROPOMI divergence data. This is necessary, since the form of the divergence output is not always circular. Furthermore, the fitted Gaussian often shows a rotation with respect to the positive north and east axes. The results of the Gaussian fits are compared to wind data in the form of a wind rose, to test the influence of the wind on the form of the divergence output, and therefore the Gaussian fit. A general case is shown in [Figure 4.16](#). The figure shows the peak-fitting sequence for a domain in Slovakia. The Gaussian is fitted on the CO divergence map (top-left). The algorithm identifies one location during 'Iteration 0' and the resulting Gaussian is depicted using contour lines (bottom-left). The fitted Gaussian is subtracted from the CO divergence map, and the residual background is shown in the top-right sub-figure. The form of the wind rose matches the output form of the Gaussian fit.

The phenomenon that the divergence output form matches the wind rose, has a simple explanation. When an emission source is emitting CO into a windy atmosphere, the emission shows a plume-like behavior. Due to inaccuracies in the TROPOMI data and the varying resolution of the data in longitudinal direction, the measured start of the plume can differ. Due to this location shift, the divergence method detects flux changes at different distances from the actual source. Since the wind has a dominant longitudinal (positive and negative) wind direction, the divergence output is elongated in the longitudinal direction. This is visible in [Figure 4.16](#). Since the Gaussian fit matches the divergence output, the fit is also elongated in the longitudinal direction. As the TROPOMI resolution only varies in longitudinal direction, this elongation of the divergence output is mostly observed in the longitudinal direction. However, due to inaccuracies in the TROPOMI data with respect to the start of the plume, the phenomenon is also observed in the latitudinal direction, albeit less frequent.

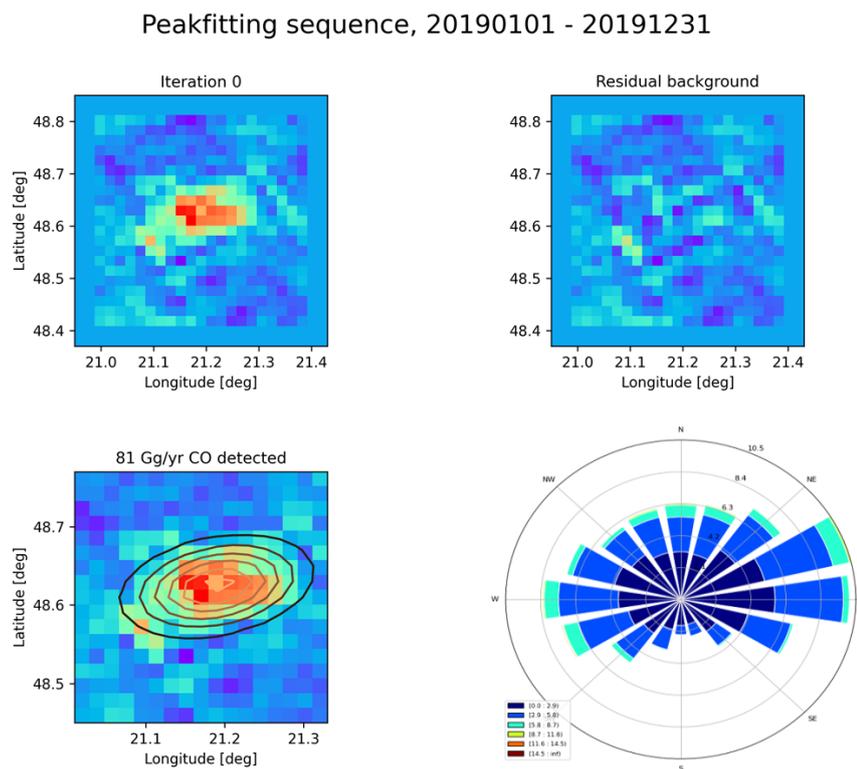


Figure 4.16: The results of the iterative peak-fitting algorithm (left bottom) is compared to the wind rose (right bottom). Similarities can be seen in the form of both outputs. Each contour line of the Gaussian fit in the sub-figure left-below represent a drop of 0.15 with respect to the highest value of the normalized divergence data.

4.4. Adaptation for the Methane Model

The different characteristics of methane as compared to CO results in the need to slightly adapt the CO model to obtain a functional model for methane. First, methane has significantly worse TROPOMI coverage compared to CO. The longer lifetime of methane results in lower relative enhancements (plume-background) and therefore only cloud-free observations are used, which results in lower coverage for methane. As stated before, coverage is an important factor in the performance of the divergence method. CO does have low coverage in scenes with high cloud coverage, but generally has sufficient coverage when analysing a full year of data. In short, if a grid is analysed for a full year of data, it mostly will yield a stable result.

The coverage of methane, however, depends largely on the geographical location. This is similar to CO, albeit that methane has close to zero coverage for some locations, as shown in Figure 4.17. In this figure, the coverage is defined as the number of TROPOMI pixels within a certain $1^\circ \times 1^\circ$ gridcell in 2019. It can be seen that especially Northern Africa, the Middle East and Australia have good coverage. Low coverage areas are South America, Siberia, China and central Africa.

When analysing a location that is located in a high coverage area and using a sufficient time interval (> 1 year), the output is expected to be stable. However, when looking at locations in low coverage areas, even with a sufficient time interval, the output will most likely be unstable. Furthermore, the TROPOMI retrieval algorithm for methane has a high dependence on aerosol optical thickness and albedo. To verify whether enhancements in the divergence/emission map are due to emission sources and not due to aerosol and albedo induced artefacts, the emission output can be compared to the albedo and aerosol optical thickness averages. Figure 4.18 shows three maps: the average albedo in the SWIR, the average aerosol optical thickness (middle) and the emission map (right).

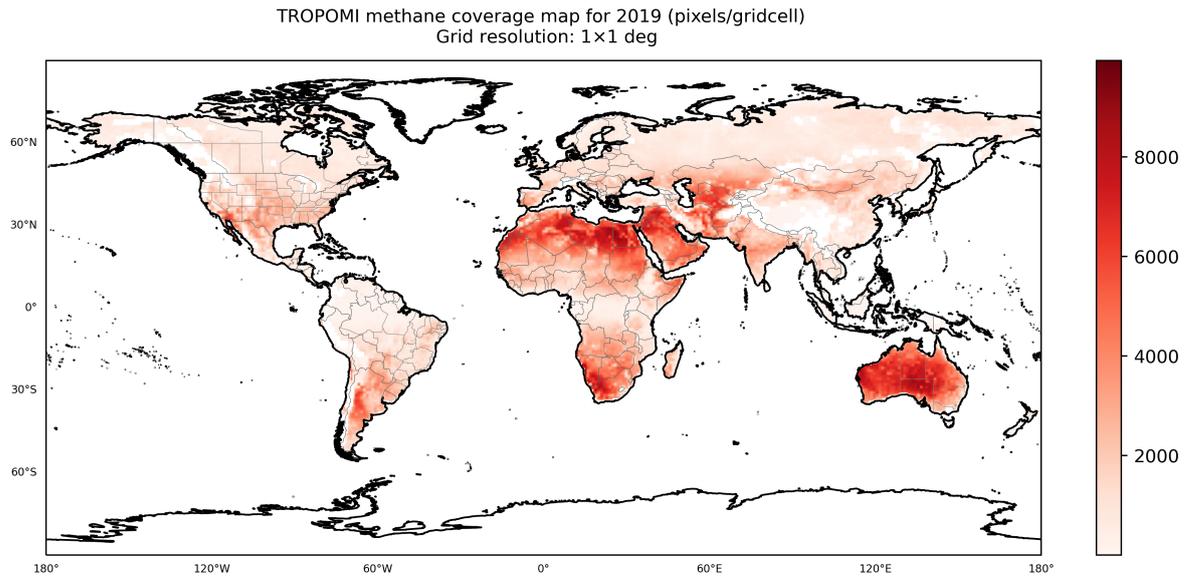


Figure 4.17: The yearly mean coverage for methane, showing high coverage in desert regions and low coverage over mountain areas.

Furthermore, the map shows locations of known coal mines in Australia (markers). The emission map shows clear enhancements around the known locations of the coal mines. However, the figure shows clear enhancements that can be related to a high gradient in the albedo map. These gradients in the aerosol and albedo maps result in a gradient in the measured methane concentration, and therefore show up as enhancements in the emission map. These enhancements are albedo and aerosol induced artefacts. Since the optimization of the methane model is not within the scope of this research, the albedo and aerosol corrections are not extensively studied, but included in the recommendations.

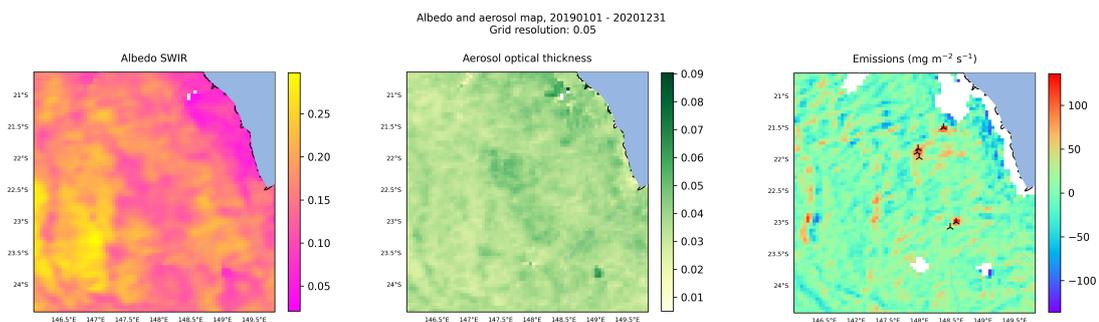


Figure 4.18: To test whether enhancements in the emission output are not artefacts, the albedo (left) and aerosol (middle) map can be compared to the emission map (right). Emission sources are identified with markers, all other divergence enhancements are artefacts.

4.5. WRF-generated CO Pseudo Data

To test the performance of the divergence method, pseudo data is used. This data is generated by the Weather Research and Forecasting (WRF)-chem CTM and the CTM is used to simulate CO concentrations around selected emission sources in Europe. The simulations use emissions estimates from the European Pollutant Release and Transfer Register (E-PRTR) for the year 2017 [70]. The

model then uses wind fields to simulate the CO concentrations in the atmosphere in the domain around the emission source on a uniform 3×3 km grid. The WRF-chem has two outputs: the gridded simulated CO concentration (xco) and the wind vectors (u, v) that are used to simulate the CO throughout the grid. The wind output is the 10 m surface wind. The simulated xco and wind variables can be used by the divergence model to test the performance of the method. To be able to use the simulation data as input for the divergence method, the TROPOMI orbits are used. To obtain a simulated CO concentration value for every TROPOMI pixel in the orbit, the weighted average of the gridded simulation data is used for every overlapping TROPOMI pixel. This way, the WRF simulations are sampled similar to as TROPOMI would have observed the data. The TROPOMI orbit files now include the WRF-simulated CO concentration and the measured TROPOMI AOT and HSL values. The WRF-chem xco data has multiple tracers. A combination of these tracers results in the full CO simulation. The tracers are defined as:

1. Background
2. Emission source
3. Public power
4. Industry
5. Other stationary combustion
6. Road transport
7. Fugitives
8. Solvents
9. Other (Shipping, aviation, waste, livestock)

Each tracer contains its own xco value, and the nine tracers are merged to obtain the total xco for a specific TROPOMI pixel. Since the tracers will be individually added to obtain the xco value, the individual tracers can be manipulated to obtain different results. For instance, the tracer (2) of a source with source strength 117 Gg/a can be multiplied by a factor 2 to obtain a source with source strength 234 Gg/a. These source strength variations can be used to test the detection and quantification limits of the model in a later stage. Furthermore, since tracers can easily be added, a tracer with an additional source can be used to test the source separation ability of the model.

The pseudo data consists of three data types: Pressure Weighted (PW), without Averaging Kernel (NO AK) and with Averaging Kernel (AK). First, the PW contains the xco value as simulated by the WRF-chem model. It has perfect coverage, and therefore can be used to test the performance of the model in an ideal situation. The dataset without AK uses the pressure weighted xco value, but also uses the TROPOMI filtering. Therefore, this set can be used to test the influence of lower coverage on the performance of the model. Lastly, the dataset with AK is the most realistic dataset, as it is the closest match to the actual TROPOMI data. The AK dataset corrects for measurements above clouds. The vertical sensitivity of the retrieval for the total column of CO is described by the AK. The AK is dependent on the cloudiness of the scene, as shown in [Figure 4.19](#). For strict cloud clearing of the data (black line), the AK is close to one for all altitudes. Using a slightly less strict data filter (yellow) to obtain the clear-sky like scenes results in a slightly reduced sensitivity for CO towards the surface. However, the sensitivity for this filter setting is reasonable. The presence of thick clouds significantly influences the vertical sensitivity of the retrieval. Due to cloud shielding, the sensitivity below the cloud is significantly reduced. This results in the retrieval estimating the CO total column

based mainly on the measurement above the cloud [71]. Therefore, it is necessary to use appropriate TROPOMI filtering, with at least a HSL (z) of $z < 5$ km and an AOT (τ) of $\tau < 0.5$. The AK dataset can be used to test the influence of the characteristics of the TROPOMI data (coverage and sensitivity) on the performance of the model.

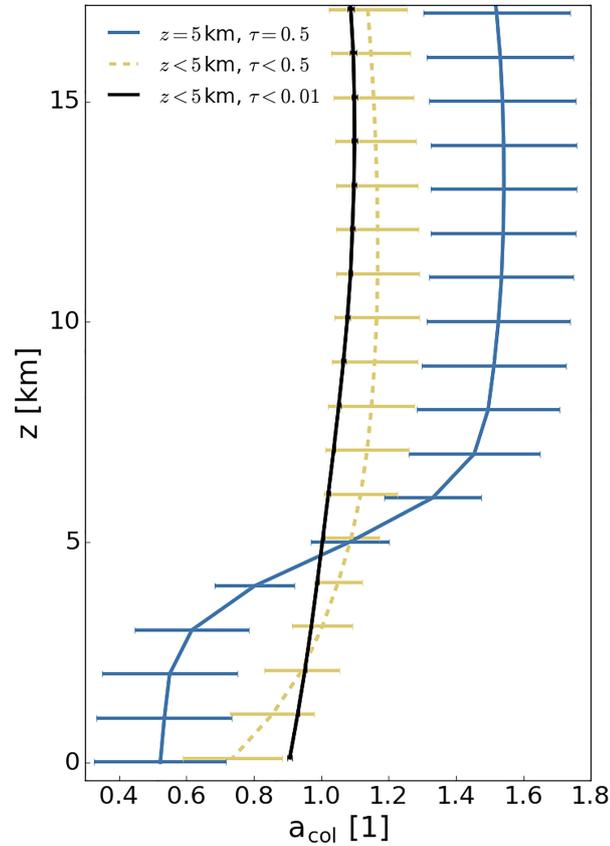


Figure 4.19: The averaging kernel for three different categories of cloudiness: strict cloud clearing (black), semi clear-sky (yellow) and high optical thick clouds (blue). The standard deviation is indicated as error bars [71].

The pseudo data contains two types of domains, named WRF locations. First, there are 18 domains containing only one source. These domains can be used to test the performance of the emission quantification of the divergence method. Second, there are two domains with multiple sources, of which one contains two sources and one contains three sources. These domains can be used to test the performance of the source separation of the divergence method.

The domains with a single source can be used to test the performance of the emission quantification. The performance is quantified by comparing the output of the Gaussian peak-fitting algorithm with the actual WRF emissions. The performance factor (P_f) is set as the relative deviation from the actual WRF emissions, and can be written as:

$$P_f = \frac{Q_{div} - Q_{WRF}}{Q_{WRF}} \cdot 100\% \quad (4.19)$$

where Q_{div} is the emission quantification of the divergence model and Q_{WRF} is the actual WRF emission quantity. To improve the overall performance of the model, a sensitivity analysis is used. This analysis varies parameter setting for filters used in the divergence model to obtain the settings with the best performance. The best performing setting is the analysis with the lowest absolute performance factor P_f .

TROPOMI Data Filtering

The first step is determining the appropriate filtering for the TROPOMI data. As previously elaborated, the sensitivity to the emission sources of TROPOMI measurements with high cloud coverage is low. One problem with data filtering for the divergence method is, however, the coverage issues. The TROPOMI data needs to be appropriately filtered without creating low-coverage issues. Therefore, multiple filter settings are tested. The least strict settings are taken as the minimal required filtering: $z < 5$ km and $\tau < 0.5$, as defined by previous research. To test whether the performance of the method improves for stricter filter settings, a sensitivity analysis is used. This analysis combines all filter settings for the height scattering layer and aerosol optical thickness: $z = [0.5, 2.5, 5.0]$ and $\tau = [0.4, 0.45, 0.5]$. This results in nine sensitivity tests. The performance of the filter thresholds are expressed in terms of the grid coverage and the performance factor (P_f). It is expected that stricter filter settings will result in a better sensitivity to emissions, but with lower coverage. The sensitivity analysis will determine if the improved estimation compensates for the lower grid coverage.

4.5.1. Sensitivity Analysis

The divergence model parameter optimization consists of separate sensitivity analyses to obtain optimal model settings for (1) dataset, (2) wind fields, (3) TROPOMI resolution, and (4) divergence model resolution. Furthermore, the pseudo data is used to test the quantification limit and the uncertainty, as well as the effect of alternative numerical differentiation methods. Each analysis calculates the performance factor for each individual sensitivity test, after which results can be compared to determine which test has the best performance. The best performing model settings are selected and used as main settings for the method, resulting in an optimized divergence model. All analyses are run on 0.04° model resolution, with the exception of the sensitivity analysis for model resolution.

Dataset

The dataset sensitivity analysis looks at the performance of the three different datasets for all WRF locations. It uses the WRF-model wind to test the influence of the dataset on the performance of the model. It is expected that the pressure weighted is the best performing model. The analysis with the AK is closest to the actual TROPOMI data, which can be used to estimate the performance of the TROPOMI data as input for the divergence method. The pressure weighted has perfect coverage, while the other two datasets use the filtering for AOT and HSL as determined by the sensitivity analysis for the TROPOMI filtering. The dataset analysis is mainly carried out to understand the influence of the AK dataset, and therefore the TROPOMI data, on the output of the divergence method.

Wind Fields

The sensitivity analysis for the wind uses three different wind fields as input: (1) GEOS-FP 10 m, (2) ERA5 10 m and (3) ERA5 100 m. The analysis is carried out with the PW dataset. This way, only the influence of the wind is tested. The analysis uses seven model runs. The analysis model runs with the following wind inputs:

1. WRF-model wind
2. GEOS-FP 10 m
3. ERA5 10 m
4. ERA5 100 m
5. GEOS-FP 10 m with wind speed filter (WSP)

6. Lowest value of GEOS-FP 10 m and ERA5 10 m
7. Highest value of GEOS-FP 10 m and ERA5 10 m

The first model run uses the WRF-model wind. This ‘perfect wind’ should result in the best performance. All other performance factors should be compared to the performance factor of the first run. The wind sensitivity analysis has three modified runs. Run (5) uses a wind speed filter, in which all values below 1 m/s and above 10 m/s are filtered out [1]. This is due to the inaccuracy of the wind measurements for wind speeds below 1 m/s. Furthermore, strong winds with speeds over 10 m/s might result in erroneous values in the divergence, as shown in Figure 4.20. The figure shows two problems with large wind speeds. First of all, the plumes are stretched out and broken up. This results in the divergence map showing a line of dipoles, instead of a positive enhancement at the beginning of the plume. A dipole is a phenomenon where a strong positive enhancement is directly bordered by a strong negative enhancement. These broken up plumes result in erroneous strong positive and negative values in the divergence, which decrease the accuracy of the method. Furthermore, due to the high wind speeds, even small variations in CO concentrations are detected by the divergence method. This results in large noise fields in the output.

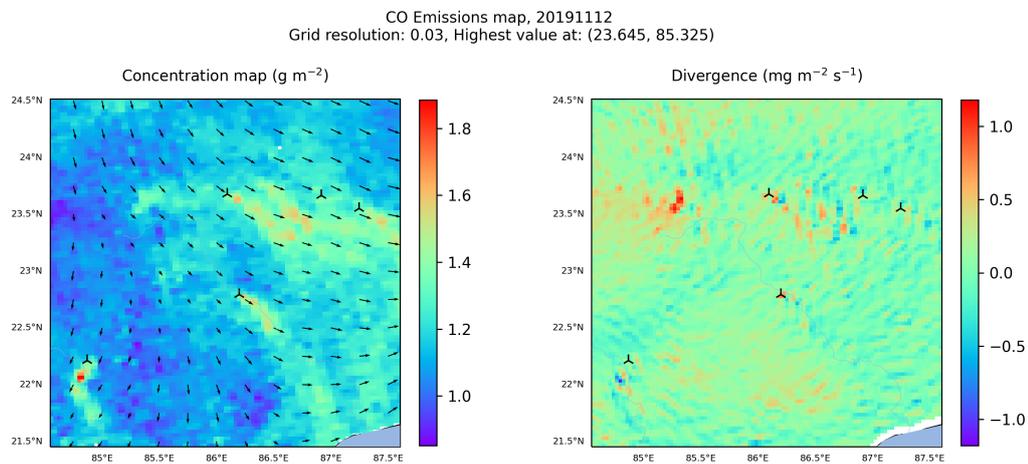


Figure 4.20: A day in 2019 with wind speeds up to 18 m/s, resulting in strong wind-induced noise fields in the divergence map. Due to strong winds, the plumes in the grid are broken-up, resulting in dipoles in the divergence.

Run (6) uses the lowest value for each gridpoint of GEOS-FP 10 m and ERA5 10 m. This is done to minimize the error due to erroneous wind speeds. However, this decreases the divergence signal strength. Run (7) uses the opposite, and takes the highest value of both wind fields. This is done to maximize the signal input for the divergence calculations. However, it is likely that this test leads to increasing noise in the grid.

TROPOMI Resolution

The test for the TROPOMI resolution is based on the varying across-track resolution of TROPOMI. As the latitudinal resolution has no variation, a sensitivity test is used only for the longitudinal resolution. The test is undertaken to investigate whether the results of the divergence method improve when the TROPOMI pixels at the edges of the swath are filtered out. For this analysis, the AK dataset with the WRF-model wind is used. Since the performance of the divergence method depends on the TROPOMI resolution, low resolution pixels can decrease the performance of the method. To

determine the filter thresholds, the TROPOMI pixel resolution for a domain Hebei, China has been used. In Figure 4.21, the across-track pixel resolution is shown. The figure contains all TROPOMI measurements in a $2^\circ \times 2^\circ$ domain with multiple orbits. The resolution varies between 0.0589° and 0.41° , with most values between 0.06° and 0.15° . Looking at July 6th, the variation of the TROPOMI pixel resolution in a single day can clearly be seen. To only filter out the most extreme outliers, two filter thresholds are chosen: $T_{res} < 0.3^\circ$ and $T_{res} < 0.25^\circ$. The Hebei domain is taken as an example and TROPOMI resolution varies per domain, dependent on the latitude.

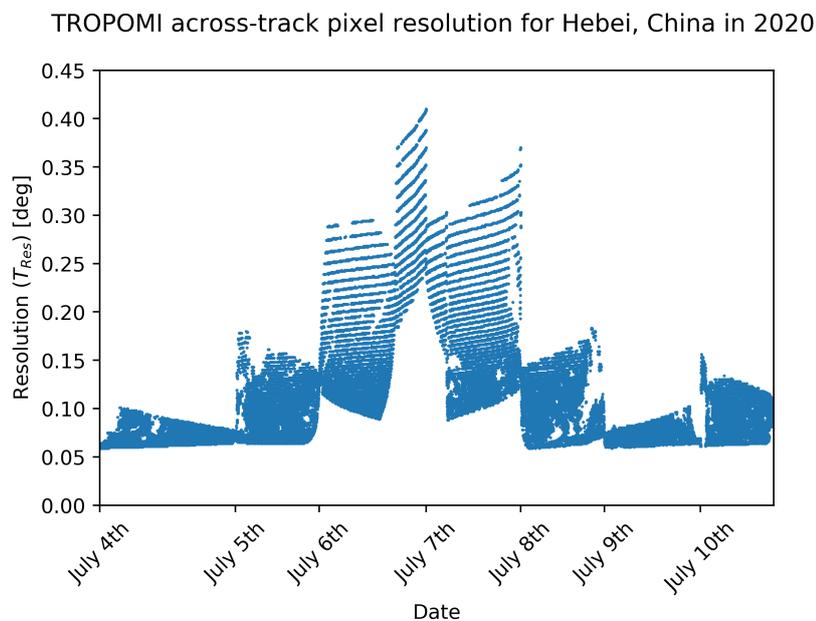


Figure 4.21: The TROPOMI across-track pixel resolution for a domain in Hebei, China, early July 2020. For each day, the large variation of TROPOMI resolution throughout the swath can be seen between 0.059° (min) and 0.41° (max).

Model Resolution

Lastly, the model resolution is tested. For this analysis, the AK dataset with the WRF-model wind is used. The model resolution both has influence on the emission quantification, as well as the source separation ability. To test source separation, a domain with multiple sources is used. Six different model resolutions are tested: $M_{res} = [0.02, 0.03, 0.04, 0.05, 0.06, 0.07]$. The divergence method aims to separate sources by using the TROPOMI data. It will not be beneficial to use model resolution lower than the TROPOMI resolution. Therefore, a maximum resolution of 0.07° is chosen. To test the highest possible model resolution, a model run on 0.01° is used. The output (Figure 4.22) shows two main issues. First, since the model uses a sub-pixel oversampling routine, the input signal is smoothed. On a resolution like 0.01° , this results in only the edges of the pixel being smoothed. Therefore, the output shows excessive stripe-like noise.

Furthermore, since the model resolution (0.01°) is about seven times higher than the TROPOMI resolution ($\approx 0.07^\circ$), it is possible that the five-point numerical method (fourth order) is located fully within one TROPOMI pixel. That means that for certain gridpoints, the divergence will be identified as zero. When a plume is detected by the divergence method, the divergence shows large positive enhancements around the start of the plume. The divergence will be slightly negative along the direction of the plume. When using too high resolutions, however, the start of the plume will show a large positive enhancement, alternated with zero values for the divergence. This causes noise in the output, and decreases the signal strength around the source. Therefore, a maximum model

resolution of 0.02° is chosen. The sensitivity analysis is used to determine whether this resolution is feasible, or too high as well. To ensure that the model does not use a too high resolution for the pseudo data analysis, a standard resolution of 0.04° is chosen for all sensitivity analyses.

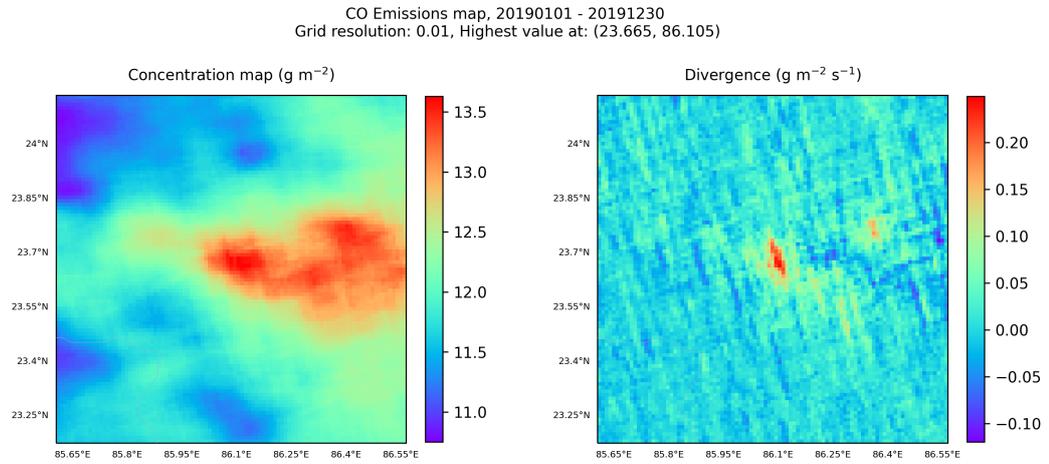


Figure 4.22: A divergence analysis for a location in India on a extremely high model resolution (0.01°). The model output shows excessive stripe-like noise.

Quantification Limit

Since the WRF-generated pseudo data consists of multiple tracers where the emission sources are all identified as a separate tracer, the source strength of the sources can be manipulated by using a multiplication factor for the specific tracer. Using this scaling, the same source can be duplicated with different emission rates and the performance of the model can be tested for different source strengths. This way, the quantification limit can be tested. The quantification limit is defined as the source strength at which the divergence method obtains a successful fit from the Gaussian peak-fitting algorithm. To test the quantification limit, a large source (> 100 Gg/a) is taken. The multiplication factors are taken as [0.05, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.75, 0.8, 0.9, 1.0, 1.5, 2.5, 5, 7.5], which results in a sensitivity analysis with 16 different source strengths ranging from approximately 5 Gg/a to 750 Gg/a (dependent on original source strength).

Uncertainty Estimation

The WRF-generated pseudo data is used to estimate the uncertainty of the model, in combination with the optimal filter settings that resulted from the sensitivity analysis. This is done by comparing the divergence method estimate with the actual WRF emissions. The relative deviation from the WRF emissions (%) is used. For each individual WRF location, this deviation is calculated. The mean (\bar{x}) of the deviation values is then calculated as:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.20)$$

where n is the number of analysed WRF locations. The standard deviation (σ) of the sample is calculated as

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)}} \quad (4.21)$$

The uncertainty of the method is now taken as one standard deviation of the sample.

Alternative Numerical Differentiation Methods

As stated before, the performance of the divergence method is directly related to the coverage of the TROPOMI CO data. High TROPOMI coverage generally results in a stable output and a successful Gaussian fit for the fitting algorithm. One disadvantage of the divergence method is the need to calculate the first-order derivative of the flux by using a numerical differentiation method. These methods enlarge data gaps and results in data loss near the borders of the grid. To test the influence of different numerical differentiation methods on the grid coverage and the performance of the divergence method, a sensitivity test is used.

After all optimal parameter values for the filter are determined by the sensitivity analysis, the sensitivity test for the numerical methods is carried out. This test uses the optimal filter setting, in combination with the pseudo data AK dataset. Two sensitivity tests are carried out, one for the second-order numerical differentiation method (Equation 4.22), and one for the fourth-order numerical differentiation method (Equation 4.23). The second order method will result in less data loss due to the 2-point formula. However, it will be less accurate than the fourth-order method. In turn, the fourth-order method will have a higher accuracy, but will result in a larger data loss. The results from both sensitivity tests can be compared to estimate whether the increase in coverage can compensate the decrease in accuracy.

$$f'(x) = \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} \quad (4.22)$$

$$f'(x) = \frac{-f(x - 2\Delta x) + 8f(x - \Delta x) - 8f(x + \Delta x) + f(x + 2\Delta x)}{12\Delta x} \quad (4.23)$$

5

Results

This chapter shows results from the divergence method. It starts with the results from the WRF-generated pseudo-data. It shows the sensitivity analyses for the TROPOMI filtering and the model filters and parameters. Furthermore, it determines the quantification limit and the uncertainty of the divergence method. Lastly, it shows seven case studies. In these studies, results from the divergence method for a specified location are compared to other quantification methods and bottom-up inventories, as well as previous research.

5.1. Optimizing the Model Parameters

As previously described in the methodology in [chapter 4](#), the sensitivity analysis for the pseudo data consists of two parts. First, the appropriate filtering is chosen for the TROPOMI data. This test uses various filter thresholds for AOT and HSL. After the filter thresholds are chosen, the sensitivity analyses elaborated in [chapter 4](#) are carried out.

For the sensitivity analyses, nine non-coastal locations from the WRF pseudo-data are used. The locations are identified with a number (identifier), as shown in [Table 5.1](#). The WRF locations include a large variety of yearly emissions, ranging from 20.5 Gg/a to 117 Gg/a. Since the sensitivity analyses are time intensive, only the Galati, Watenstedt and Duisburg domains are used for the AOT and HSL filter thresholds analysis. For the remaining analyses, all nine locations will be used.

Table 5.1: The WRF pseudo-data locations that are used for the sensitivity analyses.

City	Country	Latitude	Longitude	Emissions [Gg/a]
Galati	Romania	45.44	27.98	62.5
Gent	Belgium	51.18	3.81	94.2
Kosice	Slovakia	48.61	21.19	114
Dunaujvaros	Hungary	46.94	18.94	20.5
Saarlouis	Germany	49.35	6.75	101
Eisenhüttenstadt	Germany	52.16	14.61	38.8
Watenstedt	Germany	52.15	10.40	75
Bremen	Germany	53.12	8.68	52.1
Duisburg	Germany	51.50	6.73	117

5.1.1. TROPOMI Filtering Thresholds

The values for the filtering thresholds for AOT and HSL are chosen based on two performance parameters: the average grid coverage and the emissions quantification by the Gaussian peak-fitting algorithm. Especially the filtering on AOT results in significant data loss, as shown in Figure 5.1. Below AOT= 0.1, approximately 98% of the data is filtered out. For AOT= 0.5, this is 75%. However, this value will differ for each domain. Locations with low cloud coverage year-round, like desert regions, will see a lower data loss due to strict AOT filter thresholds compared to locations in Scandinavian countries, which have significant higher cloud coverage throughout the year.

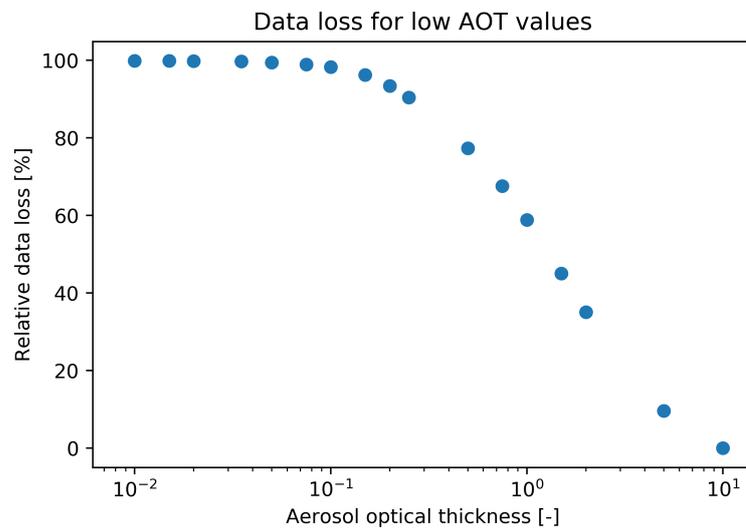


Figure 5.1: Average data loss (%) for specified thresholds for AOT filtering in Gent, Belgium.

First, the average grid coverage is analysed. The coverage is taken as the average number of data entries each gridpoint obtained over the full time domain. The analysis uses one year of data, with the possibility of multiple orbits per day, and therefore has a maximum grid coverage of approximately 480 (pixels/gridpoint). Figure 5.2 shows a 94% data loss at the most strict AOT and HSL filter settings. The least strict threshold, AOT=0.5 and HSL=5 km sees an average coverage of 76 (pixels/gridpoint). This translates to an 84% data loss.

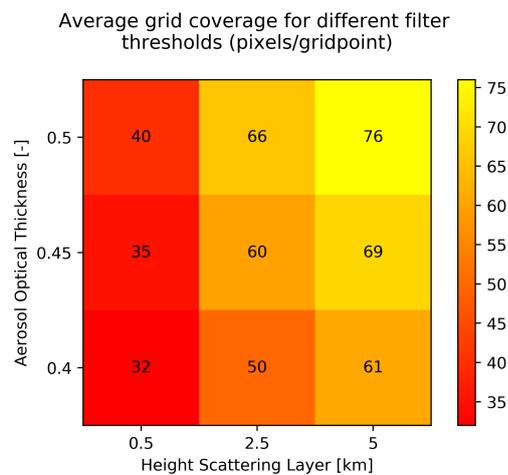


Figure 5.2: The average grid coverage for various AOT and HSL filter settings for all WRF locations, showing increased data loss for stricter filter settings.

The differences in coverage between the different thresholds are therefore significant. To determine which setting is optimal, the emission estimates must be analysed. Ideally, the most strict filter thresholds are used, but the stability of the output must be taken into account. Instability in the output means that the data does not converge to the optimal solution due to low coverage, as previously described in the methodology.

When looking at the emission estimate deviations from the actual WRF emissions for different filter thresholds (Figure 5.3), no clear improvement can be seen for stricter filter settings. Duisburg does show improvements at HSL=5 km for stricter AOT filter thresholds. However, for the same thresholds, Galati sees a degradation in emission estimate accuracy. Furthermore, some quantifications are impossible due to low coverage, which explains some missing values in the plot. This is mostly observed for AOT=0.4. For AOT=0.5 and HSL=5 km, the model performs reasonably well, with estimations within 20% of the actual WRF-generated emissions. When using stricter settings, the estimate should theoretically improve. However, due to low coverage, the Gaussian peak-fitting algorithm results become unstable, and therefore result in a worse estimation output. Since the divergence model performs best with the highest possible coverage and the data has sufficient sensitivity for all filter thresholds, the least strict settings (AOT=0.5, HSL=5 km) are chosen. These filters will be used for the remaining sensitivity analyses.

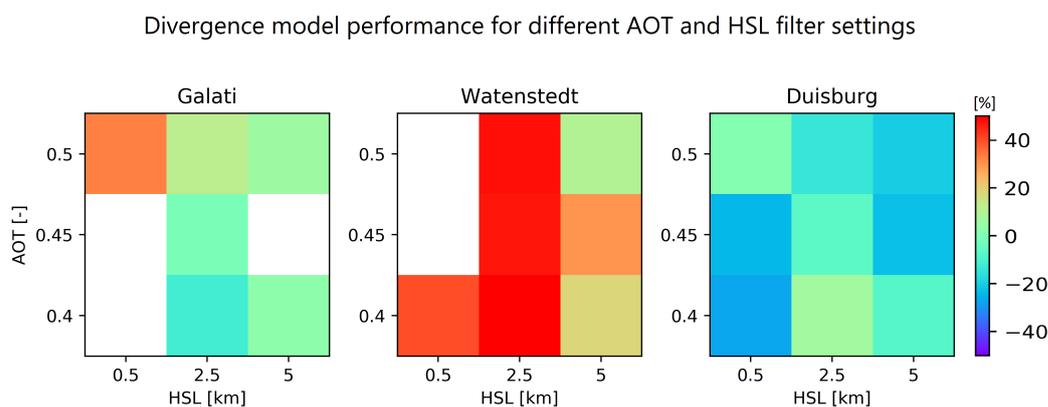


Figure 5.3: Emission estimates from the divergence method compared to the actual WRF emissions [%] for specified thresholds for the AOT and HSL filtering. No clear improvement for stricter filter settings is observed.

5.1.2. Sensitivity Analyses

For all sensitivity analyses, the two locations with the lowest yearly emissions (Dunaujvaros at 20.5 Gg/a and Eisenhüttenstadt at 38.8 Gg/a) are plotted with different markers (triangular), since these locations are below the threshold of 50 Gg/a for reasonable Gaussian peak-fitting performance. These low-emission locations can easily distort the average quantification performance.

TROPOMI Sampling

The sensitivity analysis uses three different datasets: PW, without AK (NO AK) and with AK. This analysis uses the WRF-model wind field and therefore should have minimal wind-induced error. The results for the dataset analysis are shown in Figure 5.4. From the figure it can be seen that the PW dataset performance well, with an average quantification within 15% of the actual WRF-generated emissions. The NO AK dataset has a larger spread, possibly due to the worse coverage for this dataset. The AK dataset has a good performance, with the exception of two outliers at +50% and -43%. Keeping in mind that the highest outlier is one of the low-emission locations, the performance

of the averaging kernel dataset with the previously determined TROPOMI filtering for AOT and HSL is said to be good. In general, lower coverage decreases the accuracy of the quantification. The lower sensitivity of the NO AK and AK datasets compared to the PW dataset results in a generally lower quantification. The AK dataset will be used for the coming sensitivity analyses, since it is the closest resemblance to the TROPOMI data.

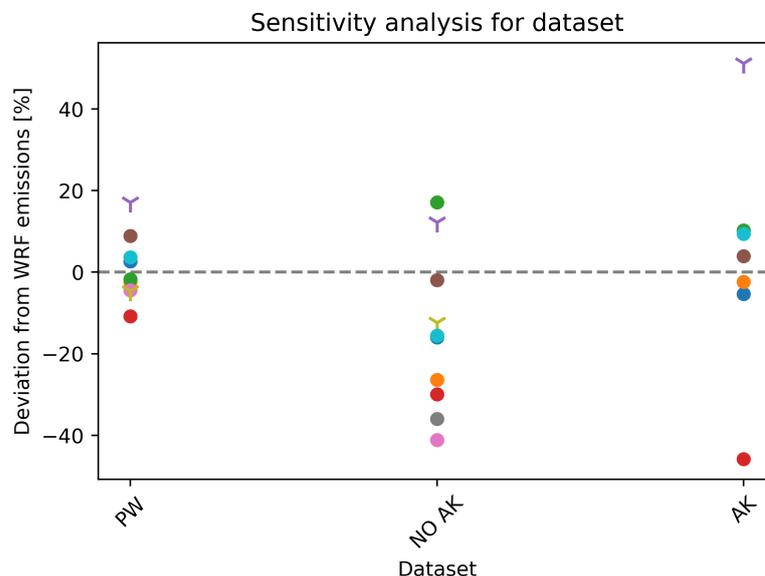


Figure 5.4: The sensitivity analysis of the 9 locations (marked by colors) for the datasets, showing the best performance for PW. The AK dataset is closest to the TROPOMI data. Sources above 50 Gg/a are marked by dots, sources below 50 Gg/a are marked by a triangle.

Wind Field

The sensitivity analysis for the wind field input is carried out with the PW dataset to only observe the effect of different wind fields on the output. All results are compared to the PW with WRF-model wind, as shown in Figure 5.5. The GEOS-FP 10 m wind leads to an average underestimation of the emissions. The ERA5 10 m wind leads to a strong underestimation. The ERA5 100 m wind has a similar wind direction as the ERA5 10 m wind, albeit with generally a higher wind speed. Therefore, the quantification estimate is more accurate than the one using the ERA5 10 m wind. Furthermore, the wind speed filter (Wind Speed filter (WSP)) improves the GEOS-FP 10 m estimate. The minimum value of the GEOS-FP 10 m and ERA5 10 m leads to a similar estimate as the ERA5 10 m wind. The maximum value of the previously named wind fields leads to a large spread in quantification and generally a large overestimation. The combination of min and max from the GEOS-FP and ERA5 10 m winds are not feasible and will not be used any further in this research. The best performing wind field is the ERA5 100 m wind.

As stated in the previous section, the wind speed filters improves the emission estimation for the GEOS-FP 10 m wind. However, this filter is not used on the ERA5 10 m and 100 m wind. Therefore, an additional analysis is carried out to test the performance of this filter on the previously mentioned wind fields, as shown in Figure 5.6. It can be seen that the wind speed filter has worse performance for the ERA5 10 m wind, possibly due to the amount of data that is filtered out since the wind field includes many wind data points below 1 m/s. The wind speed filter improves the estimation for both the GEOS-FP 10 m and the ERA5 100 m wind fields. Since the ERA5 100 m wind field was already identified as the best performing, it is chosen as the main wind field input and will be further used in the coming sensitivity analyses.

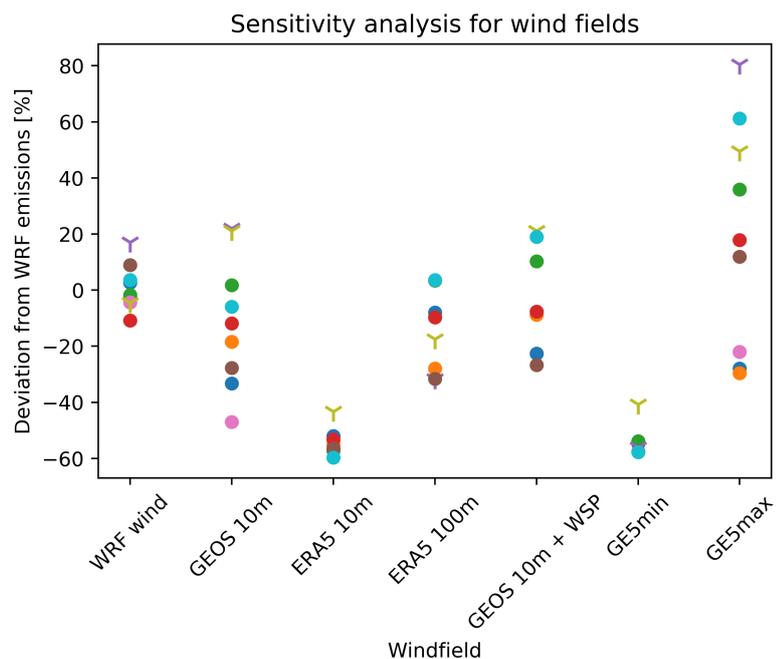


Figure 5.5: Analysis results for varying wind field input, showing GEOS 10 m and ERA5 10 m and 100 m. A combination of GEOS and ERA 5 10 m winds is shown by GE5, selecting the min or max value of both wind fields. A wind speed filter (WSP) slightly improves the GEOS 10 m estimate. The best performing input is chosen as ERA5 100 m.

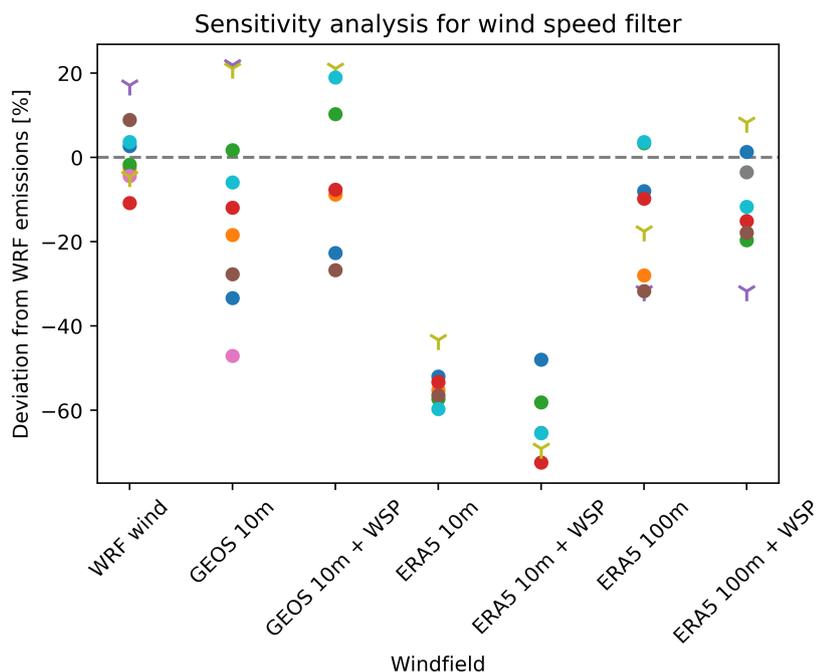


Figure 5.6: The additional sensitivity test for the wind speed filter for all individual wind fields.

TROPOMI Resolution

To test whether the TROPOMI resolution has an influence on the emission estimate, a resolution filter is used. For this analysis, the AK dataset with the WRF-model wind is used. As can be seen in [Figure 5.7](#), both filter setting (Res < 0.3 and Res < 0.25) do not result in a clear improvement of

the emission estimate. Due to the excessive data filtering, the divergence output becomes unstable and the emission estimate becomes less reliable. Therefore, it is determined not to use a TROPOMI resolution filter.

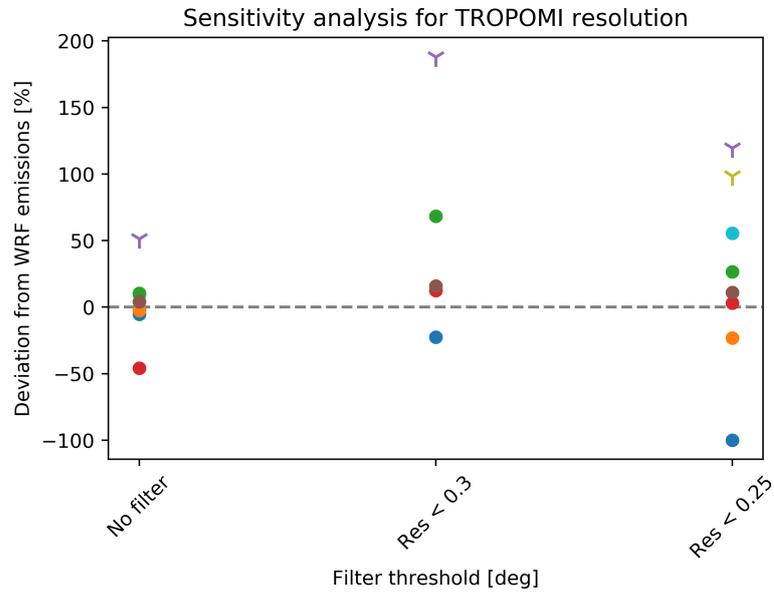


Figure 5.7: The analysis for the TROPOMI resolution filter, showing no clear improvement for both filter settings.

Model Resolution

The model resolution has influence on two factors: the ability to separate closely-spaced point sources and the emission estimation. First, only the emission estimate is taken into account, of which the results are shown in [Figure 5.8](#).

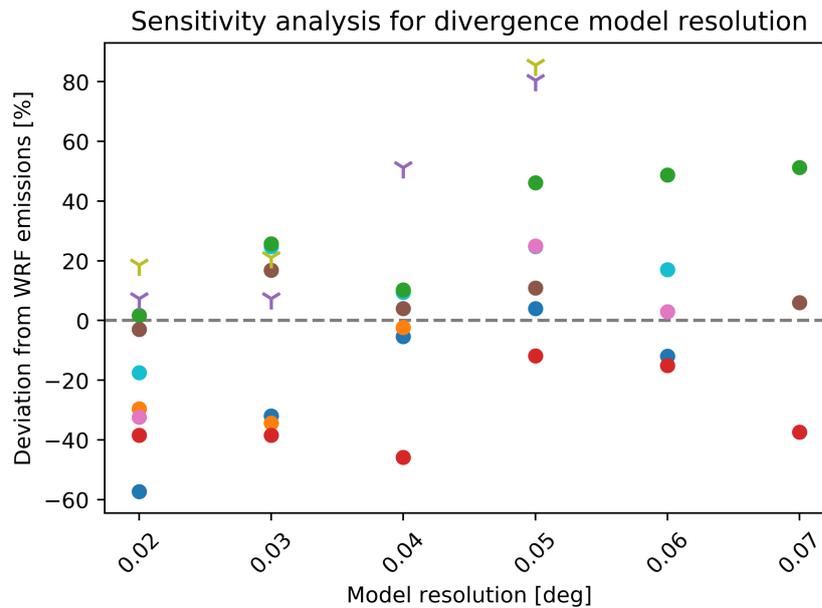


Figure 5.8: Relative performance for varying model resolutions. The best performing resolution for emission quantification is chosen as 0.04° . For resolutions $> 0.06^\circ$, not all model analyses resulted in a successful fit, which makes that these resolutions show less results.

For the highest resolution of 0.02° , the quantification is largely underestimated. This can be caused by excessive noise due to a high model resolution, as previously shown in [Figure 4.22](#) on page 50. Furthermore, due to the high model resolution, the emission output shows less Gaussian behavior, which results in an inaccurate fit. For lower model resolutions, the quantification accuracy improves, with an optimum at 0.04° . When using lower resolutions (0.06° and 0.07°), some source signals are not suitable for the Gaussian peak-fitting algorithm, since the divergence enhancements are limited to only a few gridpoints. Therefore, the optimal resolution is 0.04° , with reasonable performance for both 0.03° and 0.05° .

5.1.3. Multiple Source Domains

The model resolution directly influences the ability to separate closely-spaced point sources. For the separation test, Duisburg is used. This single-source location can be combined with additional source domains, which consists of the same grid as Duisburg, but with a different source. Since multiple domains are included in the domain, the sources are individually identified and are shown in [Table 5.2](#).

Table 5.2: Additional sources for the multi-source domain sensitivity analysis.

Plant name	Latitude	Longitude	Country	Emissions [Gg/a]
Thyssenkrupp North	51.50	6.73	Germany	117
Thyssenkrupp South	51.49	6.71	Germany	7.3
Hüttenwerke	51.37	6.72	Germany	185

The Thyssenkrupp sources (N/S) are only 3.5 km apart. This makes it unlikely that the sources can be separated by the divergence method, since both sources will often fall within the same TROPOMI pixel. Furthermore, the difference in source strength is significant, which makes it even more unlikely that the sources can be separated. Therefore, the Thyssenkrupp South plant is enhanced by a factor 10, to make the sources more equal in strength and to make sure the source is above the quantification limit. The Hüttenwerke plant is located approximately 13 km from the northern sources. This makes it possible to separate the sources, since the sources will often fall within different TROPOMI pixels.

The Gaussian peak-fitting sequence is run for different model resolutions, starting at 0.02° up to 0.07° . For the highest resolution (0.02°), the source centers are approximately 7 gridpoints apart. For the lowest resolution (0.07°), this separation distance is only 2 gridpoints. The fitting results are shown in [Figure 5.9](#). The figure shows four sequences for different resolutions: 0.02° , 0.03° , 0.04° and 0.07° . The fitting sequence starts at 'Iteration 0'. This is the initial divergence output. The algorithm fits a Gaussian (shown by the contours of the fit) in 'Iteration 1'. The fitted Gaussian is subtracted from the divergence map and the residual is used for 'Iteration 2', in which again a Gaussian is fitted (shown by the contours). For both the model resolutions of 0.02° and 0.03° , the algorithm successfully identifies both sources. However, for 0.04° , the algorithm identifies the signal as one source, and 'Iteration 2' is not undertaken, since no further sources are identified by the algorithm. For 0.07° , the sources can hardly be separated by eye, and the algorithm also is not able to separate the sources. The two northern sources are not separated for any of the model runs.

As stated before, the optimal model resolution for emission quantification was found to be 0.04° , with reasonable performance for 0.03° and 0.05° . However, for the source separation, it is desirable to use the highest possible resolution of 0.02° , as shown in [Figure 5.9](#). Since a trade-off between source separation and emission quantification performance is necessary, the optimal model resolution is chosen as 0.03° . This resolution has reasonable performance for both source separation and emission quantification.

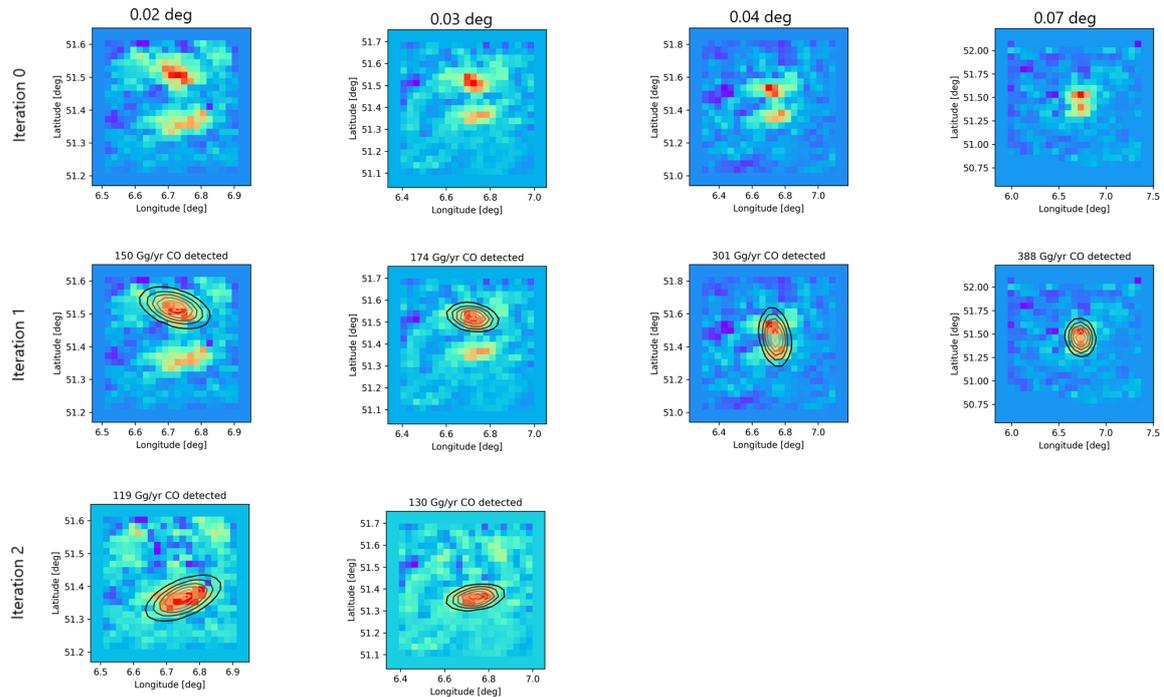


Figure 5.9: The Gaussian peak-fitting sequence for different resolutions. Iteration 0 shows the initial divergence output. Iteration 1 shows the first peak-fitting procedure. Iteration 2 shows the second peak-fitting procedure. The sources can be separated up to 0.03 deg. Each contour line of the Gaussian fits in the subfigures represent a drop of 0.15 with respect to the highest value of the normalized divergence data.

5.1.4. Uncertainty Estimation

To estimate the uncertainty of the method, the deviation from the actual WRF emissions for all nine locations is used, as shown in Figure 5.10. This analysis uses the ideal model set-up as determined in the previous sections. The figure shows the deviation from the actual WRF emissions (%) for the different source strength of the WRF locations. There is no correlation observed between the deviation from the WRF emissions and the source strength. The uncertainty of the model is taken as the mean absolute error of the pseudo data analysis. However, since the sample size is only $n=9$, the estimation cannot be determined reliable. However, since no additional WRF locations are available, this dataset will be used for the calculation.

The mean absolute error of the dataset is calculated as

$$\epsilon = 20.11 \approx 20 \quad (5.1)$$

Therefore, an uncertainty of 20% will be used for the analysis.

5.1.5. Detection and Quantification Limit

The WRF-generated pseudo data can also be used to test the detection and quantification limits of the divergence method. The quantification limit analysis uses the AK dataset with WRF-model wind. First, the detection limit in terms of source strength (Gg/a) can be determined as the ability of the Gaussian peak-fitting algorithm to have a successful fit. Second of all, the quantification limit is the ability of the algorithm to obtain an accurate fit and can be tested for various source strengths. The fit is said to be accurate when the quantification value lies within the uncertainty limit of the actual WRF-quantification. For this analysis, the Thyssenkrupp North plant (Duisburg) is used, af-

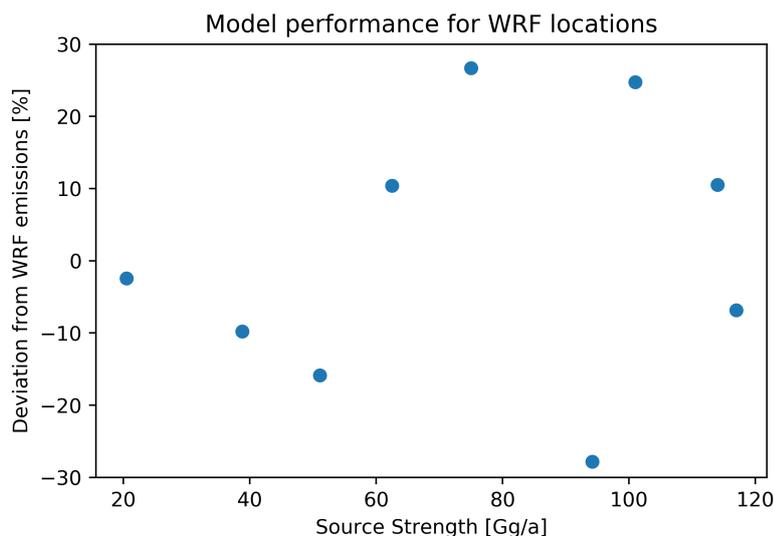


Figure 5.10: Relative model performance for the emission estimation for all nine pseudo data locations.

ter which the source strength is varied for different analysis runs. The lowest source strength is taken as 5.85 Gg/a, the highest is taken as 877.5 Gg/a. The results are shown in Figure 5.11. Two source strength runs did not have a successful Gaussian fit: 5.85 Gg/a and 11.7 Gg/a. The detection limit is observed to be 23.4 Gg/a, as this is the weakest source with a successful Gaussian fit. For simplicity purposes, this detection limit is set as 25 Gg/a. For the weakest sources (< 40 Gg/a), the algorithm overestimates the emission quantification. The results converge to 20% underestimation of the actual emissions. All source strengths above 50 Gg/a are observed to fall within the [-20%, 20%] quantification uncertainty interval. Therefore, the quantification limit for reasonable Gaussian peak-fitting performance is taken as 50 Gg/a, since it stays within limits of the model uncertainty of the converging value.

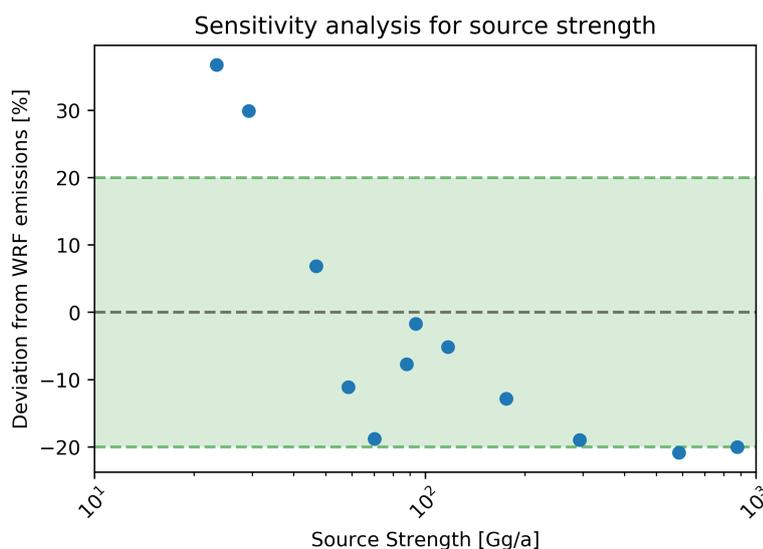


Figure 5.11: Model performance for varying Source Strength (SS). The results converge after SS = 300 Gg/a. The results are within the bounds of uncertainty after 50 Gg/a.

5.1.6. Alternative Numerical Differentiation Methods

The results of the numerical differentiation method sensitivity tests are shown in Figure 5.12. The fourth-order method performs better in terms of quantification estimates for 7 out of 9 locations. The Saarlouis and Duisburg (Thyssenkrupp North) domains yield better results for the second-order method.

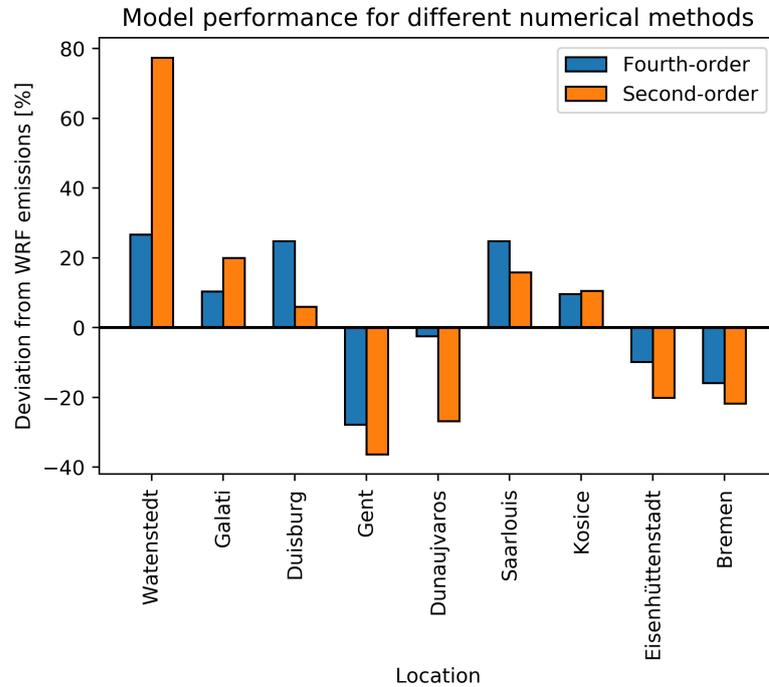


Figure 5.12: Relative model performance for the emission estimation for all nine pseudo data locations for a second-order and fourth-order numerical method. The fourth-order outperforms the second-order method for 7 out of 9 locations.

The performance of both methods can be expressed in terms of the mean and mean absolute error of the performance factors (P_f) for all nine WRF locations. The results are shown in Table 5.3. The mean and error are both calculated with respect to the actual WRF-emissions. The coverage is taken as the average grid coverage in number of pixels per gridcell.

Table 5.3: Results for the sensitivity tests of the two numerical methods.

Method	Mean (μ) [-]	Error (ϵ) [-]	Coverage [pixels/gridcell]
Second-order	2.7	34.7	79
Fourth-order	4.5	20.1	65.9

Averaged over all nine WRF locations, the grid coverage of the second-order method is 20% higher compared to the fourth-order method. Both methods tend to slightly overestimate the emission. The mean absolute error for the second-order method is 76% higher than the error of the fourth-order method. This is a clear indication that the second-order is less accurate than the fourth-order method.

The second-order method can be applied in case studies where there is no successful Gaussian fit due to coverage issues for the fourth-order method. The second-order method can increase the coverage, albeit with lower accuracy. However, if no fit is possible when using the fourth-order method, a less accurate - but successful - fit, might be desirable.

5.2. Case Studies

Using the divergence method, optimized in [section 5.1](#), six cases are extensively studied for TROPOMI CO data. The locations are chosen such that they vary in geographical location, seasonal coverage and source type and strength to test the model performance in varying environments. All locations have multiple closely spaced sources, which vary from a distance of 3 km to 50 km. The case studies use three years of CO data (01/July/2018 - 30/June/2021) for the model test. The case studies are sorted in order of increasing complexity, dependent on the number of emission sources, altitude variation, coverage and type of analysis. An additional location is studied using TROPOMI methane data. This domain is not extensively studied, but only used to test the functionality of the methane divergence model. The case study locations and the main reason for including the domains in this research are shown in [Table 5.4](#).

Table 5.4: Locations for the case studies with the main reasons for including the domains in this research.

Name	Country	Lat	Lon	Main characteristics
Duisburg	Germany	51.4	6.8	Two very closely spaced steel plants, in close vicinity of large cities.
Hebei	China	39.9	118.6	No a priori information about emission sources, strong seasonal dependence for coverage.
Jharkhand	India	23.0	86.1	Multiple steel plants, large surface elevation variation, strong seasonal dependence for coverage.
Cairo	Egypt	30.1	31.2	Multiple emission sources within the city of Cairo, coverage mostly influenced by Nile delta.
Riyadh	Saudi-Arabia	24.5	46.8	Multiple sources within city limits, high coverage year-round, ability to perform a day-to-day analysis.
Wuhan	China	30.5	114.2	Multiple steel plants close to the city, ability to test influence of Covid-19 pandemic on CO emissions.
Korpezhe	Turkmenistan	38.1	54.0	Multiple oil and gas facilities, good coverage, region is extensively studied and documented.

The case study analysis consists of five steps. First, a domain is chosen. This can be done by using known locations of emission sources or analysing a larger domain (a country as a whole) on a low (1°) resolution and selecting a zoomed-in domain with large concentration enhancements. Second, the CO concentration map on a 0.07° model resolution is used to verify that the domain includes strong CO concentration enhancements and to test whether individual sources can be identified by using a model resolution similar to the TROPOMI resolution. This lower resolution (compared to the optimized model resolution) is used to decrease the model run time. Third, each location will undergo the full model analysis using the filters and resolutions specified in the methodology. The output will identify and quantify the individual point-sources in the grid. The artefact identification analysis ([subsection 4.3.3](#)) is carried out to verify whether the output of the Gaussian peak-fitting algorithm are emission sources. This can be done by looking at the coverage and/or individual days to observe plumes. Next, the unknown emission sources can be identified by using satellite imagery or databases. Lastly, emission quantification results from the divergence method can be compared to quantification analyses of previous research and bottom-up inventories such as REAS (Hebei and Jharkhand), E-PRTR (Duisburg), TNO (Duisburg), DACCIWA (Cairo) and EDGAR (all locations). A flowchart showing the full case study procedure is included in the appendix ([Figure C.1](#)). An overview of all identified sources, including emission estimates and locations, can be found in [Table C.1](#).

5.2.1. Duisburg, Germany

The Duisburg location is located in the west of Germany, close to the Dutch border. The location is characterized by the river Rhine and has no significant elevation. The grid size is taken as $1.2^\circ \times 1.2^\circ$, centered around the city of Duisburg. This specific domain is chosen since it includes three known steel plants around the city of Duisburg. As shown in Figure 5.13, the area has large CO concentration enhancements. In the figure, city contours of several large cities (> 500,000 inhabitants) are shown. Since no correlation between urban density and CO concentration is observed, it is concluded that all CO concentration enhancements are caused by industry.

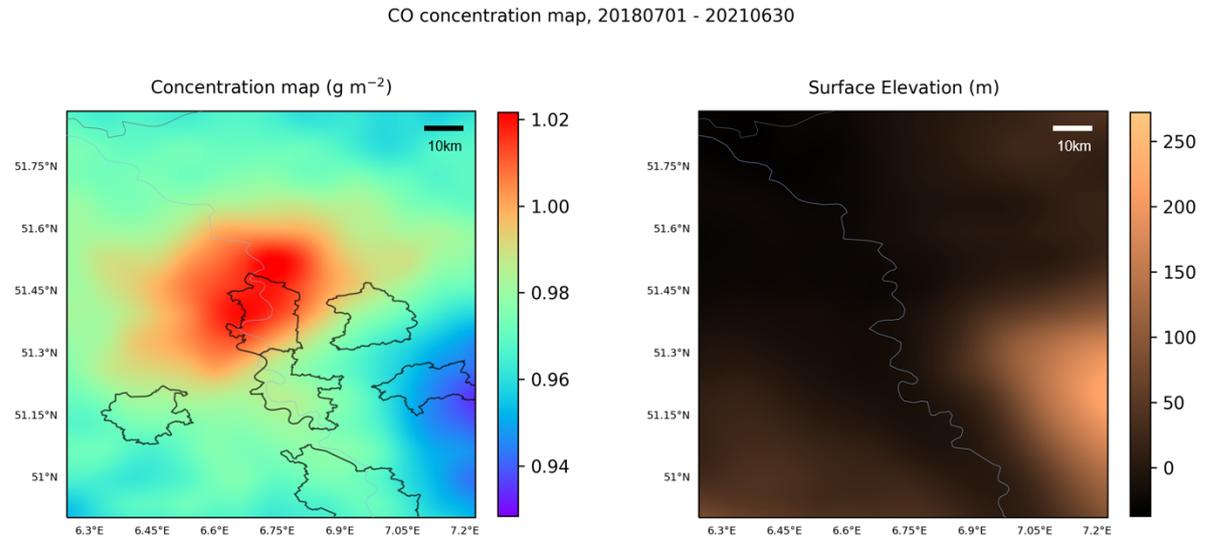


Figure 5.13: The Duisburg area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°), including contours of large cities. Large CO concentration enhancements in the center of the grid can be seen.

The coordinates and emission estimates of the known steel plants are shown in the table below, as taken from the E-PRTR [70]. As the coordinates indicate, both Thyssenkrupp plants are located close to each other. The distance between both plants is approximately 3.5 km. Comparing this distance to the highest TROPOMI resolution of 7×5.5 km, it is likely that both plants often fall within the same TROPOMI pixel. Therefore, it is unlikely that both sources can be separated by the divergence method. The Hüttenwerke Krupp plant is located approximately 13 km from the Thyssenkrupp plants.

Table 5.5: Known sources in the Duisburg area. Three steel plants are identified [70].

Name	Latitude	Longitude	Emission estimate [Gg/a]
Thyssenkrupp North	51.50	6.74	117
Thssenkrupp South	51.48	6.71	7.3
Hüttenwerke Krupp	51.37	6.72	185

To test whether divergence enhancements might be artefact induced, the coverage (Figure C.12) and pixel vs divergence maps can be used (Figure C.13). Both figures do not indicate any coverage issues, so it can be assumed that all divergence enhancements are source related. Since all divergence enhancements can be related to known source locations, there is no need to further study the artefact/emission source classification, as all divergence enhancements are classified as emission sources. Now, the results of the Gaussian peak-fitting algorithm are compared to the TNO inventory (Figure 5.14). The locations of the model results and the TNO inventory match nicely, albeit with a

small shift in center for the Hüttenwerke plant (as compared to the location of the marker).

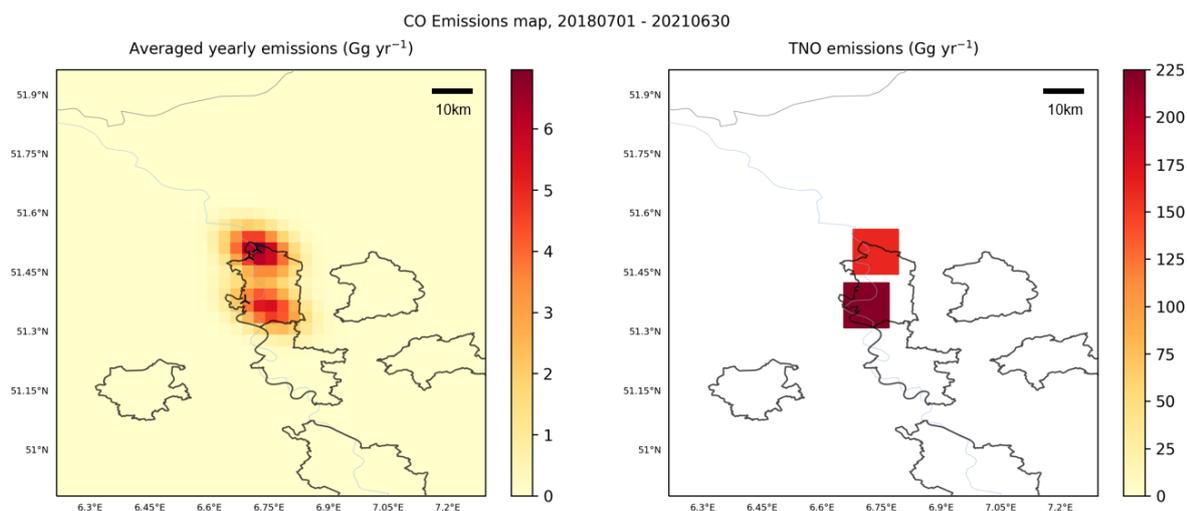


Figure 5.14: Emission estimates from the divergence model per gridcell for the two identified plants in Duisburg (left), compared to estimates from the non-gridded TNO inventory for 2017 (right).

For the Duisburg location, four quantification datasets are used: (1) the divergence method results, (2) the EDGAR inventory, (3) the TNO inventory, and (4) the quantification values as published by the E-PRTR. The results are shown in Figure 5.15. The EDGAR inventory also includes emissions from other sectors besides industry, whereas the other inventories in the figure only include emissions from steel plants.

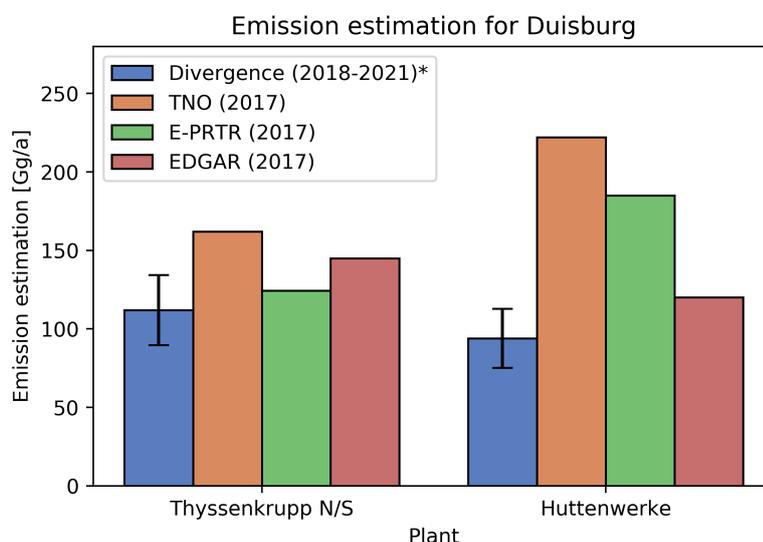


Figure 5.15: Quantification estimates for the Duisburg case study. The results of the divergence model are compared to different inventories. * Results for the divergence method are shown for 01/July/2018 - 30/June/2021.

For the Thyssenkrupp N/S plant(s), the emission estimates from the inventories are within the uncertainty of the divergence method, with the exception of the TNO inventory. The Hüttenwerke plant, however, sees a strong underestimation compared to the inventories (TNO and E-PRTR). Since this source has an underestimation compared to the inventory estimates, it is expected that the method has data loss around the southern source, which caused the center to shift and the estimate to be lower. However, the EDGAR quantification is just outside the uncertainty of the di-

vergence estimate. Since the inventory also includes urban emissions, it can be expected that the emission estimate from EDGAR for industry nicely matches the divergence results. lastly, since the emission inventories use 2017 data, it is possible that the estimates are outdated and therefore non-matching. Concluding, the divergence has good performance for the Thyssenkrupp N/S plants in terms of quantification and localization.

5.2.2. Hebei, China

The east of China lacks overall information about emission source locations and quantification. Since no a priori knowledge of source locations is present, a feasible domain cannot be determined before the analysis. Therefore, the east of China (Hebei province) is first analysed using the divergence method on a low resolution (0.5°) to identify large CO concentration enhancements. The low resolution analysing is carried out using a $10^\circ \times 10^\circ$ domain centered on (38.0, 112.0). Multiple CO concentration enhancements can be seen in the domain (Figure 5.16). The concentration enhancement around (37.0, 114.5) is located around the city of Xingtai. However, since the city of Xingtai is one of the most polluted cities in China [72], CO concentration enhancements might be the result of diffuse urban emissions. Therefore, the Xingtai location is not chosen as domain for the high resolution analysis. The concentration enhancement around (39.5, 118.5) is located in the less populated Hebei region and a domain around this location is chosen (black box, Figure 5.16) for a high resolution analysis.

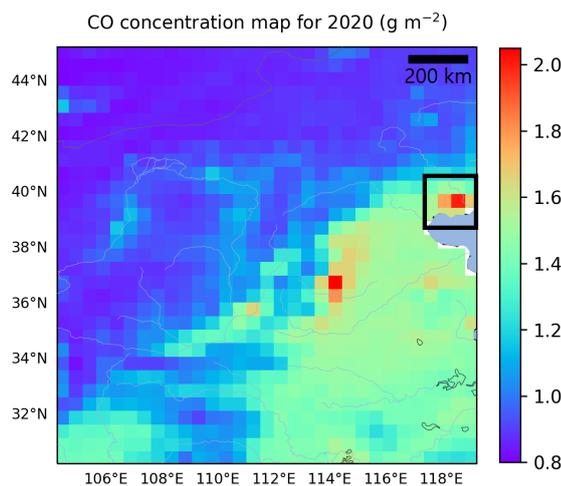


Figure 5.16: The low resolution (0.5°) analysis for East-China (2020), with two main concentration enhancements over Xingtai and Hebei. The area marked by the black box is chosen for the high resolution analysis.

The Hebei domain is located close to the capital Beijing. Since no previous research is available for the Hebei case study, there are no known emission sources in the grid. The grid size is taken as $1.2^\circ \times 1.2^\circ$. The domain has no mountain ridges throughout the grid, but sees the surface elevation increasing gradually from south to north (Figure C.7). The domain is densely populated and contains many industrial plants, including some iron and steel plants, of which the exact locations are not known. The domain borders the Chinese Yellow Sea in the south and sees the foothills of the Mongolian highlands in the north, with altitudes up to 750 meters. The domain is characterized by many rivers and lakes. The divergence model analysis shows multiple divergence enhancements in the domain, of which the strongest is observed in the center of the grid, as shown in Figure 5.17. The divergence output indicates multiple sources.

By using the wind rose (Figure C.9), a prediction can be made about the feasibility of the divergence enhancements. Since the windrose indicates that the Hebei area has a dominant north-western

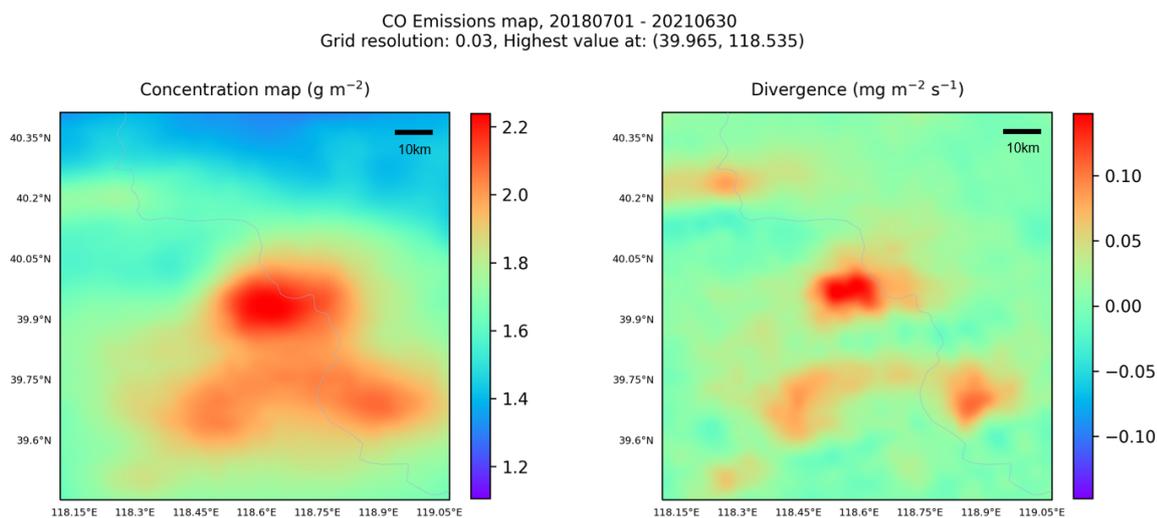


Figure 5.17: The Hebei area, showing the TROPOMI CO concentration (left) and the corresponding divergence (right) at high resolution (0.03°).

wind direction, it is expected that the CO concentration has enhancements south-east of possible sources. From Figure 5.17, it can be seen that the divergence enhancements are located in the north-west of CO concentration enhancements. Therefore, it can be assumed that the divergence output is realistic, according to the windrose.

To test whether divergence enhancements might be artefact induced, the coverage (Figure C.12) and pixel vs divergence maps can be used (Figure C.13). Both figures do not indicate any coverage issues, so it can be assumed that all divergence enhancements are source related. To further classify the output signal in source/artefact, the individual days will be observed. In Figure 5.18, two TROPOMI observations in early 2019 are shown. The figures shows at least four clear plumes in the grid. Therefore, it can be concluded that the grid contains emission sources.

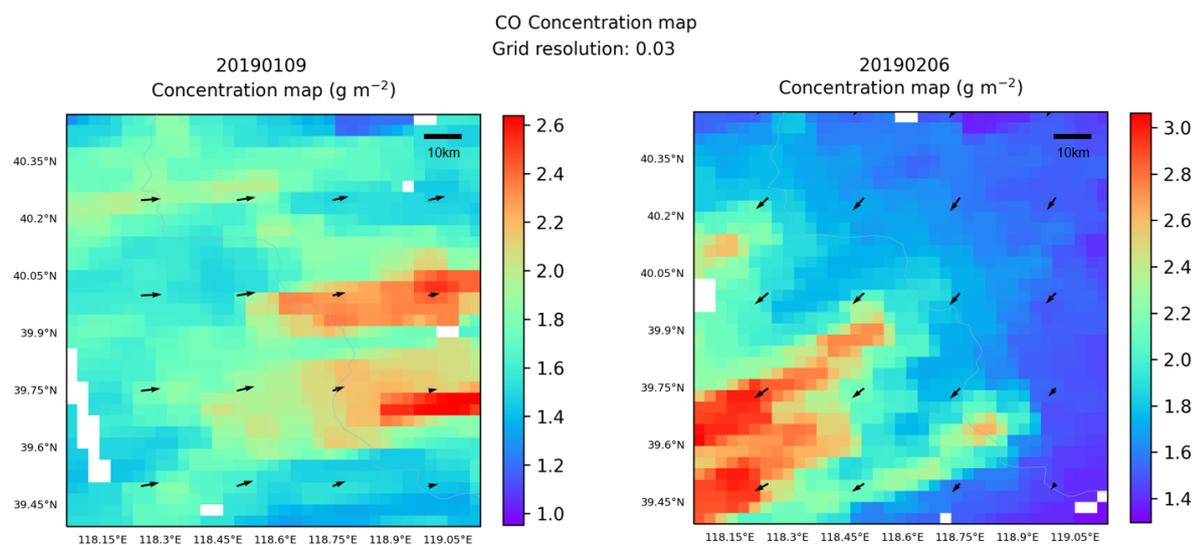


Figure 5.18: The Hebei area, showing the TROPOMI CO concentration for January 9th, 2019 and February 6th, 2019 at high resolution (0.03°). The black arrows indicate the ERA5 100 m wind direction and speed. Multiple plumes can be seen, which indicate multiple emission sources.

Next, the Gaussian peak-fitting algorithm is used to identify and quantify the emission sources in the grid. The algorithm identifies five sources, ranging from 1,400 Gg/a to 210 Gg/a (Figure 5.19).

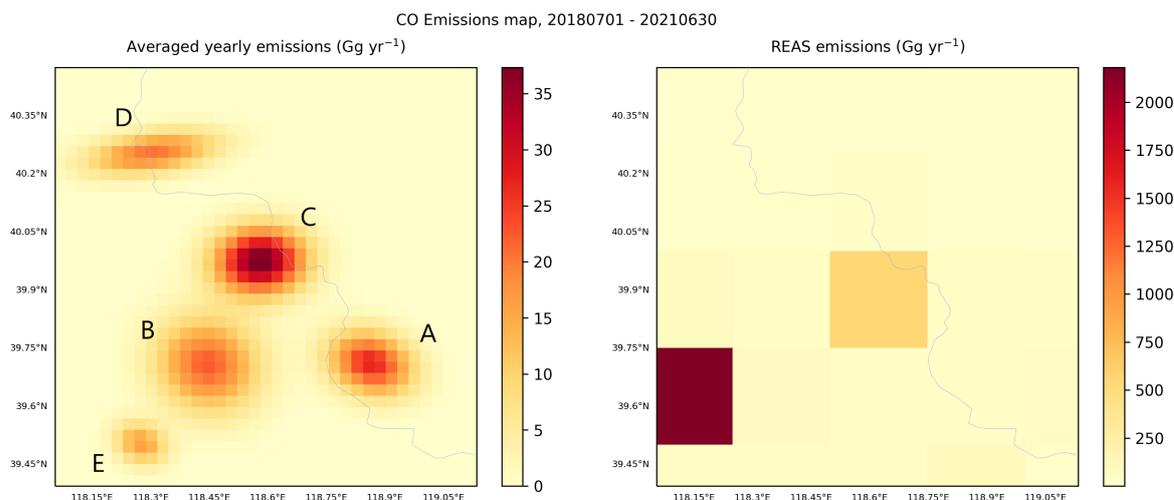


Figure 5.19: The Hebei area, showing the results from the peak-fitting algorithm (left). Five sources are identified (A-E). The emission estimates can be compared to the REAS inventory (right). It can be seen that neither the locations of the sources nor the emission estimates match. Both emission estimates (divergence method and REAS) show estimates per gridcell.

To verify that the TROPOMI observations from Figure 5.18 and the Gaussian peak-fitting algorithm (Figure 5.19) actually match emission sources, the locations are observed with satellite imagery. All five locations match a steel plants in the Hebei area (Figure C.11). The quantification estimates are shown in Table 5.6. The sources are identified using the Global Steel Plant Tracker from Global Energy Monitor (GEM) [73].

Table 5.6: Five identified sources for the Hebei area with the locations and emission estimates from the divergence method using TROPOMI CO data, including estimates for iron and steel production [73].

ID	Name	Lat	Lon	Emission estimate [Gg/a]	Crude steel [Mt/a]	Crude iron [Mt/a]
A	Qinhuangdao Hongxing	39.66	118.89	1,300 [1,000-1,500]	2.72	3.85
B	Tangshan Donghai	39.63	118.42	850 [680-1,000]	5.5	5.0
C	Qian'an Liangang Yanshan	39.92	118.68	1,400 [1,100-1,600]	10.0	10.0
D	Hebei Jinxi	40.21	118.22	620 [500-750]	5.0	7.4
E	Tangshan Ruifeng	39.47	118.26	210 [170-250]	6.245	5.817

The quantification of the model estimate can be compared to REAS and EDGAR inventories. However, this particular area is not well documented in both databases. For EDGAR, no specific point-sources are included in the inventory. For the REAS inventory only source (C) is included. This source strength is estimated to be 560 Gg/a by REAS, compared to 1,400 [1,100-1,600] Gg/a by the divergence method. Furthermore, the largest source in the south-west of the REAS grid (2,181 Gg/a) is not identified by the divergence method, but also does not show any enhancements in the CO concentration map. This reflects the lack of reliable bottom-up data over China and shows that the divergence method can add valuable information there.

The Global Steel Plant Tracker does not include estimates for emissions but does include estimates for iron and steel production, shown in megatonnes per year (Mt/a), which are based on official

reports. From [Table 5.6](#), there is no correlation observed between the emission estimates from the divergence method and the iron and steel production estimates. The Qinhuangdao Hongxing (A) plant produces approximately half the amount of iron and steel of the Tangshan Ruifeng (E) plant, but emits 6 times as much CO per year. In [Figure C.8](#), the difference is shown visually. Based on the emission estimates from the divergence method, it is expected that the production estimates from the Global Steel Plant Tracker are incorrect and/or incomplete.

The Hebei case study showed to importance of the divergence method in regions where information about emission sources (location and quantification) is scarce. The divergence method identified and quantified five large emission sources without any a priori information about these sources, underlining the independence of the method.

5.2.3. Jharkhand, India

The Jharkhand domain is centered on Bokaro Steel City (23.0, 86.1), which is located in the east of India in the Jharkhand province. The area is characterized by steel production, but also has multiple coke plants in the vicinity. In total, the Jharkhand state steel production was estimated to be 1.275 Megatonnes per year in 2017, which makes up about 12% of the total steel production of India [74]. The grid size is set as $3^\circ \times 3^\circ$ around Bokaro Steel City. With a mean elevation of 250 meters above sea level and a maximum elevation of 850 meters within the grid, there is a large variation in surface altitude. Furthermore, the orography is characterized by mountain ridges throughout the grid. Towards the east of the grid, the first signs of the west-Bengal river delta can be spotted. The rest of the grid is characterized by the Indian highlands. The grid includes two rivers: Subamarekha river in the center and Brahmani river in the south-east of the grid. The coverage has a high seasonal dependence due to the monsoon, with low coverage in June - October ([Figure C.21](#), [Appendix C.4](#)).

The low-resolution CO concentration and corresponding surface elevation maps show multiple concentration enhancements and strong elevation gradients throughout the domain ([Figure 5.20](#)). Furthermore, the concentration map shows low CO concentration values at gridcells with high surface elevation. This shows the added value of an altitude correction for the Jharkhand domain.

CO concentration map, 20180701 - 20210630

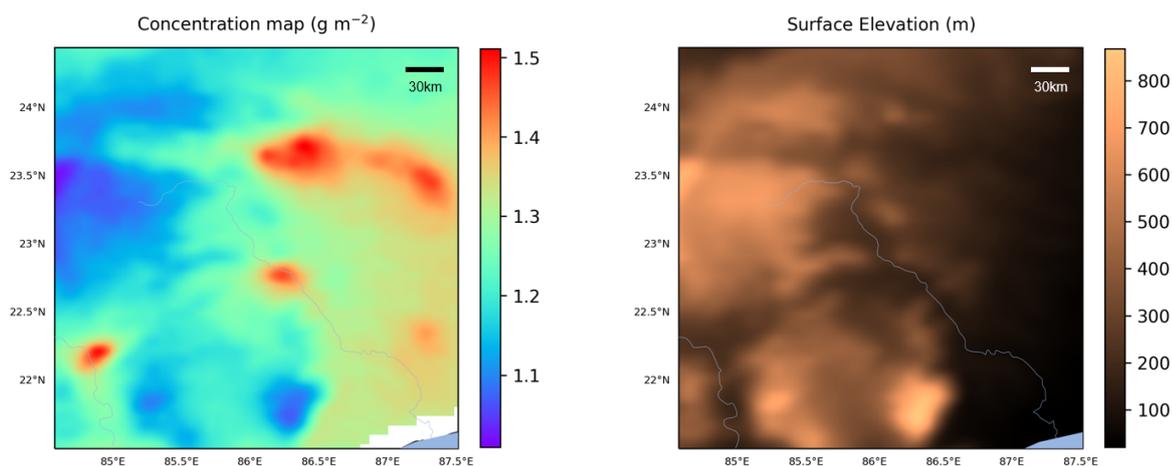


Figure 5.20: The Jharkhand area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Multiple enhancements can be spotted.

To further identify and quantify the emission sources, the divergence method is used. To analyse the performance of the source identification, the output is compared to previous research by Sadavarte and Venkataraman [75].

Table 5.7: Locations and emission estimates for five known steel plants in the Jharkhand area [75].

Name	Latitude	Longitude	Emission estimate [Gg/a]
Rourkela	22.21	84.87	299
Durgapur	23.55	87.25	249
Bokaro	23.68	86.10	600
IISCO	23.67	86.92	87
TATA	22.79	86.20	697

To conclude whether the locations of the known sources match the model results, the emission sources from Table 5.7 are plotted in the model output (Figure 5.21). In the concentration map, it can be seen that the concentration enhancements are not centered around the markers. For the Bokaro plant in the northern part of the grid, the concentration enhancement is mostly distorted towards the east. The Durgapur plant (most eastward plant of the grid) sees enhancements mostly south of the marker. Surprisingly, the Rourkela plant (south-west) has enhancements mostly towards the south-west of the plant.

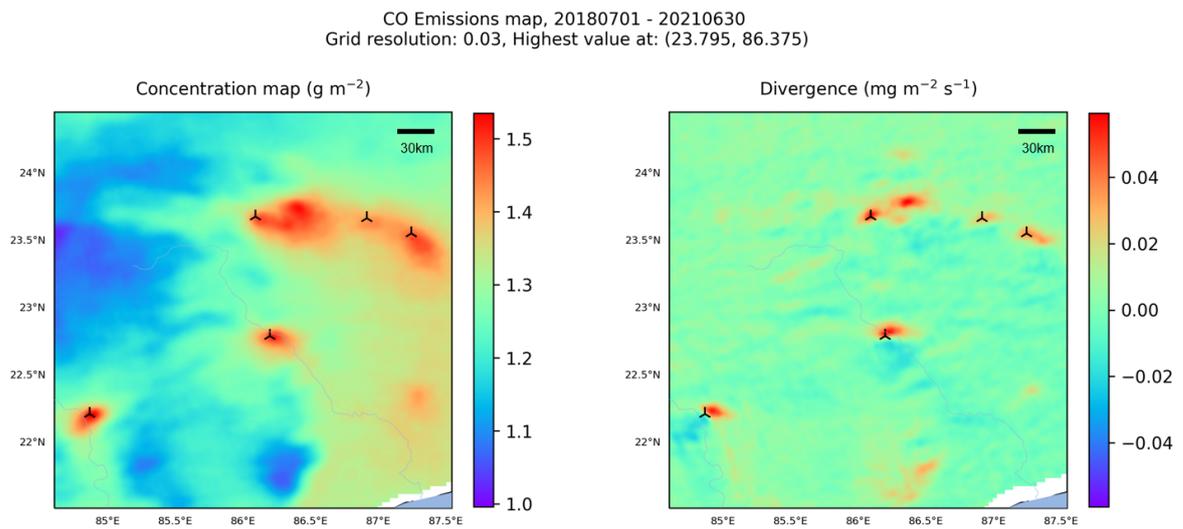


Figure 5.21: The Jharkhand area, showing the TROPOMI CO concentration (left) and corresponding divergence (right) at high resolution (0.03°). The markers indicate the locations of known emission sources.

These results can be explained by the wind rose, shown in Figure C.16 in Appendix C.4. From the windrose, it can be concluded that the full grid does not have one dominant wind direction. However, the northern part of the grid has west as the dominant wind direction, the south of the grid has south-west as the dominant wind direction and the east part of the grid has south as the dominant wind direction. This matches the transported enhancements around the emission sources, as shown in Figure 5.21. In general, if the windrose shows a dominant wind direction for a source location, it is expected that the CO concentration enhancement is located downwind of the dominant wind direction, as seen from the source location.

The divergence method identified eight sources, of which three are not previously known. Plume observation is used to verify the unknown emission sources (Figure 5.22). Plumes can be seen emerging from the known locations, as well as other source locations in the grid. The divergence enhancements perfectly match the plume locations. Therefore, all identified sources are identified as emission sources. The geographical locations of these sources are observed using satellite imagery, as shown in Figure C.18 in Appendix C.4. Source (A) corresponds to a cement/power plant, which is identified as the Metaliks cement and power plant. Source location (B) is located on top of a coal mine in the Jharkhand area. Since coal production has no clear CO emissions, it is unlikely that the coal mine is the main source of the CO emissions. Source location (C) matches the Saluja Steel plant.

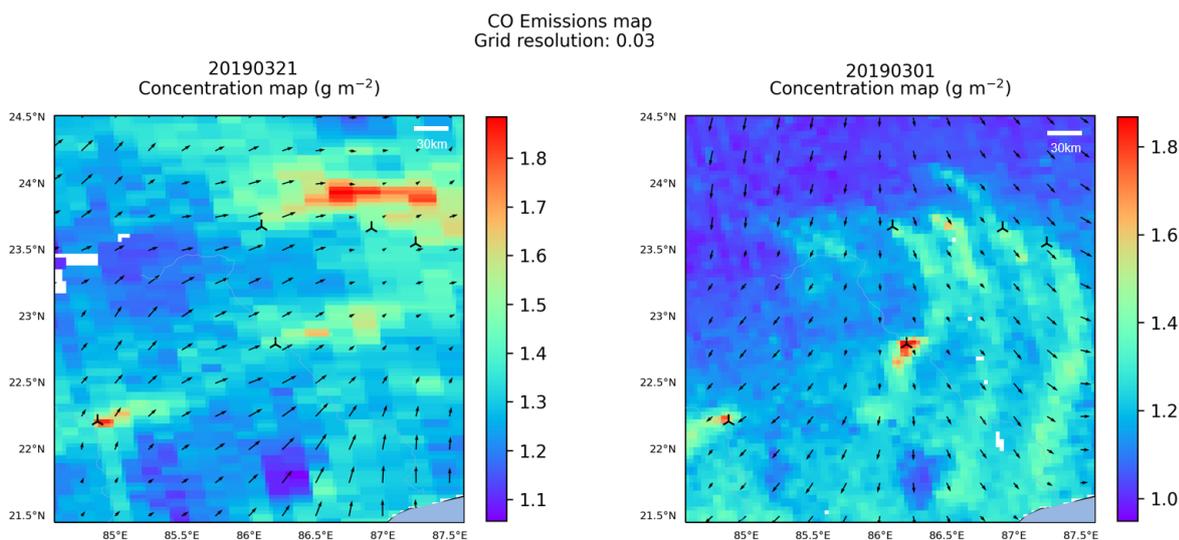


Figure 5.22: TROPOMI CO concentration at high resolution (0.03°) for March 1st and March 21st, 2019. The black arrows indicate the ERA5 100 m wind direction and speed. Multiple plumes can be seen, which indicates multiple sources. The signal from the unknown location does show plume-like behavior, which indicates a source.

Since no possible emission sources are identified as artefact, the sources can be compared to inventories (Figure 5.23). When looking at the REAS inventory, one location stands out. According to the REAS inventory, a CO emission source of 121 Gg/a is located at (23.625, 85.375). However, the model does not identify the divergence signal as a possible source. Furthermore, for the Rourkela and TATA plant, the output of the Gaussian peak-fitting algorithm does not perfectly match the marker. This might be due to the large physical size of the plants, especially the TATA steel plant, which stretches out over 16 km in north-south direction. The markers are taken as the center of the plant, and therefore do not necessarily match the emission sources of the plants.

Figure 5.24 shows that the divergence method has similar results compared to the REAS inventory. For the Rourkela plant, the REAS estimate is lower than EDGAR, the divergence method and Sadavarte and Venkataraman [75]. The TATA and Metaliks plants are hugely underestimated in the EDGAR database, compared to the other estimates. In general, the EDGAR estimates are overestimated, compared to the other estimates. Since the EDGAR emission estimates also include other emission sectors besides industry, some other anthropogenic emissions (urban, road transport) can be included in the gridcells matching the location of the studied steel plants. This might result in a higher estimate for EDGAR.

To conclude the case study, it is seen that the divergence method has good performance for both emission quantification estimations as well as source localization and separation for the Jharkhand case study. The case study also showed the importance of altitude correction.

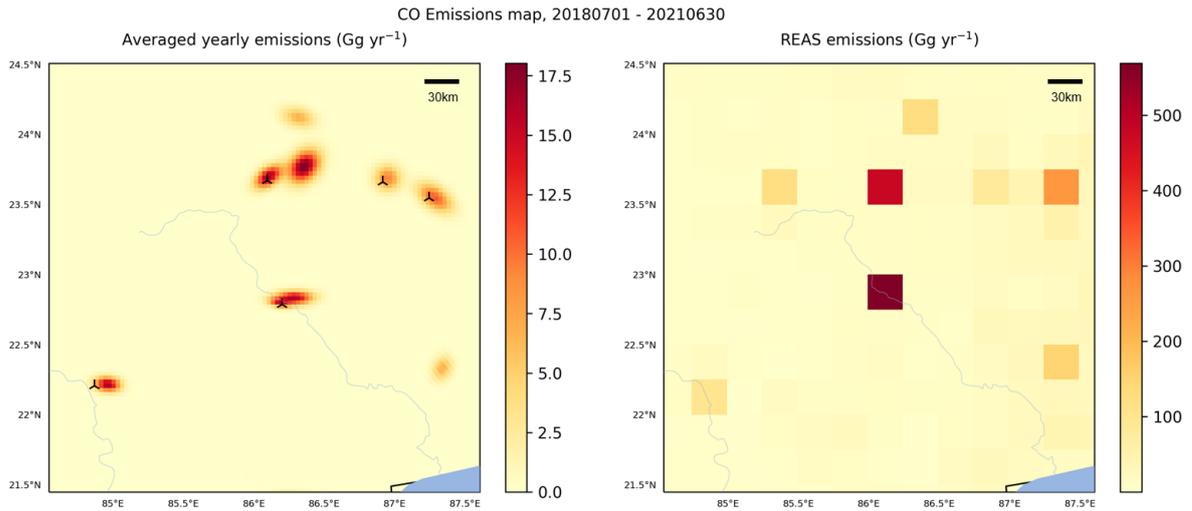


Figure 5.23: The identified emission sources by the Gaussian peak-fitting algorithm (left) and the REAS inventory (right). Three unmarked sources can be seen, which indicate unknown emission sources.

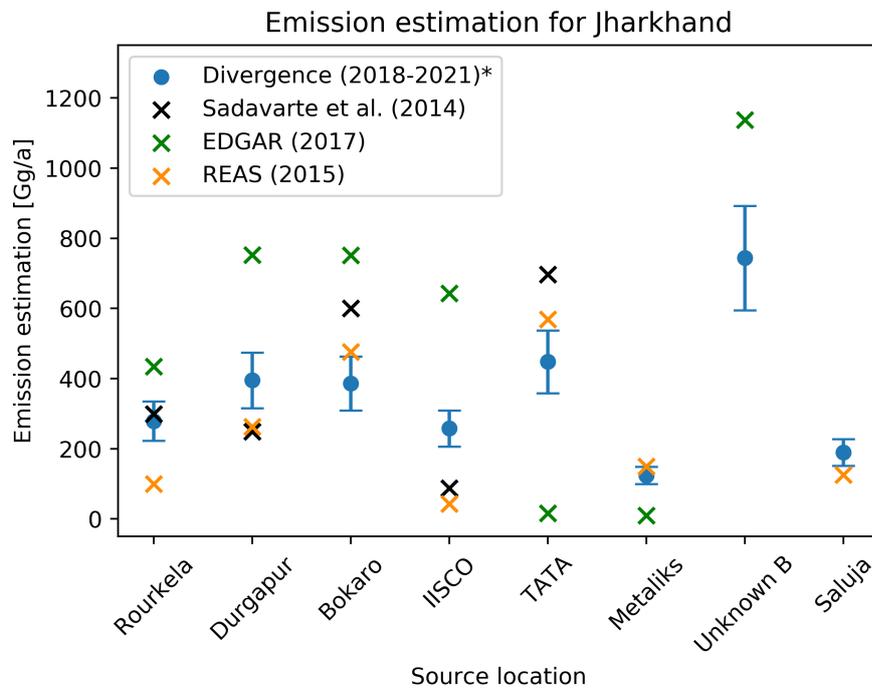


Figure 5.24: Emission quantification estimates from the divergence method, previous research by Sadavarte and Venkataraman [75] and bottom-up inventories. * Results for the divergence method are shown for 01/July/2018 - 30/June/2021.

5.2.4. Cairo, Egypt

Cairo, the capital of Egypt, is located close to the Mediterranean sea and is characterized by the Nile delta. The Nile delta can be seen in the elevation plot (Figure 5.25), starting in the north of the grid and merging into the river Nile. In the south-east of the grid, the foothills of the Suez mountains can be seen. Since the elevation difference is approximately 600 meters (between the minimum and maximum elevation), altitude correction is relevant.

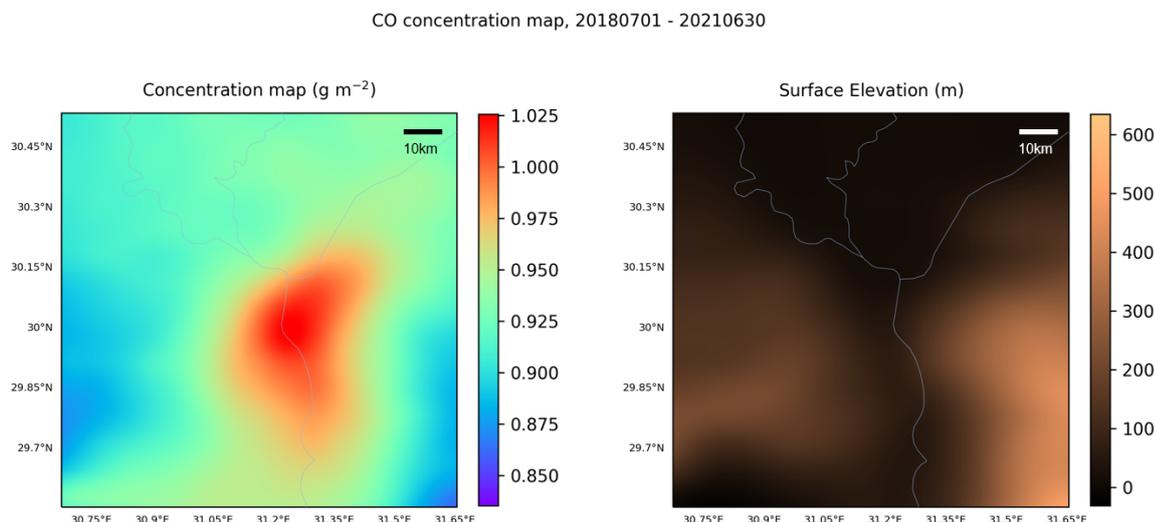


Figure 5.25: The Cairo area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Large CO concentration enhancements in the center of the grid can be seen.

The grid size is taken as $1.2^\circ \times 1.2^\circ$, centered around the city of Cairo. The city of Cairo includes multiple point-sources. However, it is unlikely that these sources will be individually visible for two reasons. Firstly, they are located close to the city. Secondly, the CO concentration enhancements from possible point-sources can not be distinguished from other sources (road transport, for instance) and the background signal coming from the city. With 9.5 million inhabitants, approximately equal in size as London, the city of Cairo is expected to have significant emissions. Urban emissions include emissions from traffic/transport, households, and industry.

When looking at the divergence output in Figure 5.26, the prediction about the separation of the point-sources with respect to the city background is confirmed. In the figure, one large enhancement can be seen along the Nile river, which is similar to the physical layout of the city. The contours of the city of Cairo are shown in the figure, but the area around the entire Nile river is densely populated.

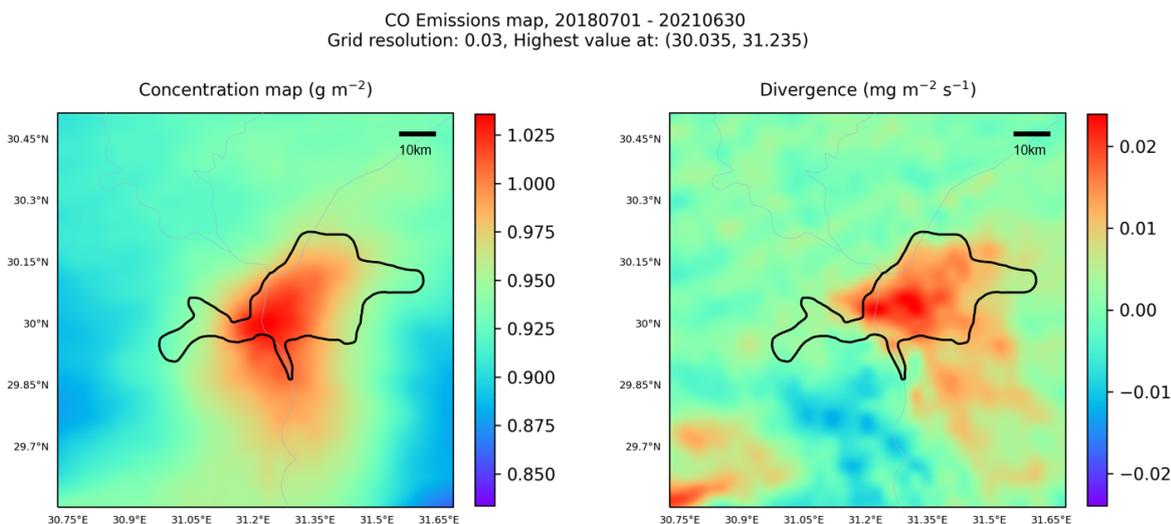


Figure 5.26: The CO concentration and corresponding divergence for the city of Cairo (outline). The city center shows large concentration and divergence enhancements. Multiple divergence enhancements can be seen east of the Nile river and in the south-west of the grid, possible altitude-correlated.

When looking at the divergence map, multiple scattered enhancement blobs can be spotted towards the south-east of the grid. These enhancements are mostly located in a mountainous desert area. Therefore, it is expected that these enhancements are altitude related and therefore can be classified as artefacts. This can be confirmed by the Gaussian peak-fitting algorithm in a later stage. To test whether the domain includes coverage-induced artefacts, the coverage maps are used (Figure C.24, Figure C.24). Extreme outliers can be seen, which indicate an artefact. Emission sources normally have a more gradual increase in divergence. Therefore, individual days should be checked to see whether the divergence enhancement can be related to an artefact.

The Gaussian peak-fitting algorithm localizes one source, which is identified as the city of Cairo, including multiple power plants in the vicinity of the city. The divergence enhancements towards the south-west and the south-east of the grid are not classified as sources, as expected. The EDGAR inventory matches the output from the divergence method around the city of Cairo. However, it shows a possible emission source in the south of the grid (23.6, 31.0) where the divergence method does not show any enhancements. Satellite imagery does show some minor industrial plants around (29.543, 30.921). Since these sources seem to be small in size, the EDGAR inventory may overestimate these sources. Furthermore, the EDGAR inventory does show gridcells with estimated emissions above the detection limit of 25 Gg/a, but no gridcells show emissions above the quantification limit of 50 Gg/a.

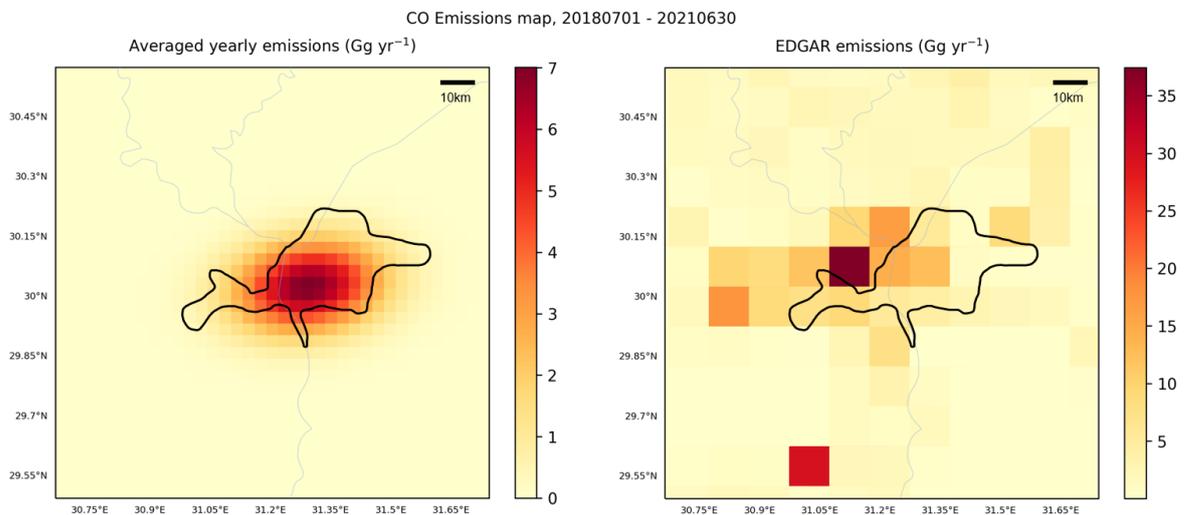


Figure 5.27: Emission estimates from the divergence method (left) and the EDGAR inventory (right). Estimates are shown as average emissions per gridcell.

The Gaussian peak-fitting algorithm does not identify the divergence enhancement in the south-west of the grid as an emission source. However, this might be due to the fact that the enhancement is located at the edge of the grid. Therefore, it is necessary to re-center the grid, and zoom in on the divergence enhancement to determine if it should be classified as artefact or not. The re-centered and zoomed in grid is shown in Figure 5.28. The divergence enhancement does not clearly indicate a point source. The stretched-out form of the enhancement is artefact-like. Point-sources generally have Gaussian shaped divergence enhancements.

To further classify the enhancement as artefact, the individual days are observed. When looking at days with different dominant wind directions, the plume signal is expected to have a different orientation, parallel to the dominant wind direction. Artefacts show similar CO concentration enhancements on days with different dominant wind directions, as can be seen in Figure 5.29. The two days with different wind directions have close-to equal CO enhancements. This indicates a clear artefact.

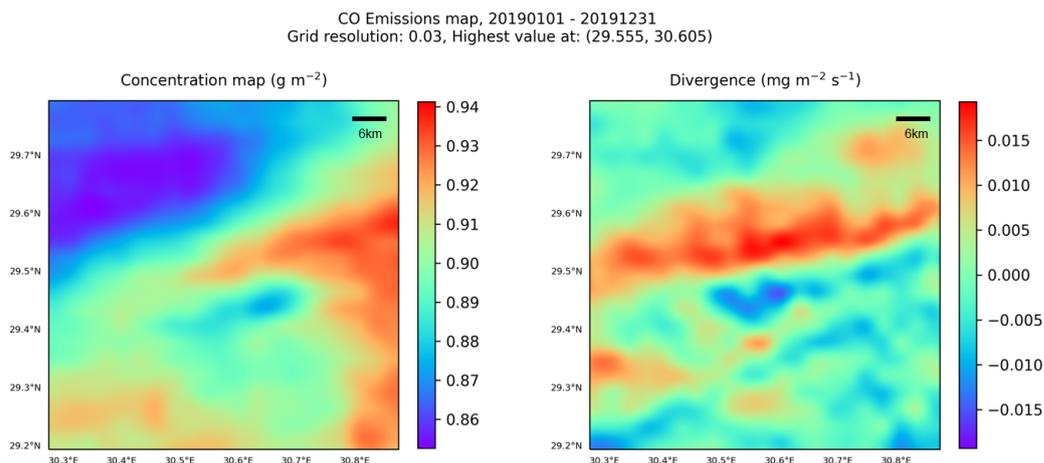


Figure 5.28: A new model run with the possible artefact location as grid center. A non-Gaussian enhancement can be seen in the divergence.

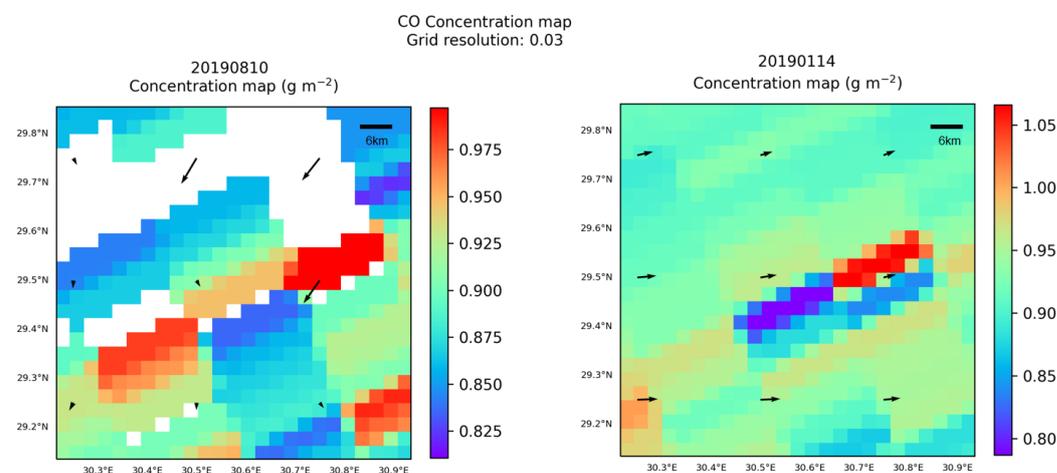


Figure 5.29: Two individual days (August 10th and January 14th, 2019) showing CO enhancements, but no plume-like behaviour.

Lastly, the location of the artefact is observed with satellite imagery, as shown in [Figure C.26](#) in [Appendix C.5](#). The location of the divergence enhancement is just above Qaraun Lake. The darkness of the water, combined with the brightness of the desert sand makes it likely that this location is an artefact. Furthermore, there are no possible emission sources (industrial plants) observed in the vicinity of the location. Therefore, it is classified as artefact.

For artefact identification, one other method can be used. When looking at the coverage map, the previous case studies only looked at the relation between high divergence enhancements and low coverage. However, the low coverage problem is not necessarily common when looking at artefacts. Since the divergence method uses a 5th percentile background correction, the artefact enhancement is generally not filtered out by the background correction, since it is expected to be among the highest values in the grid. This is confirmed in the coverage map of the artefact ([Figure C.27](#)). The high coverage enhancement in the pixelmap is the location of the artefact, which is visible due to the background correction inability to filter out the artefact. Since emission sources have plumes that move with the wind, the coverage map will show enhancements in various directions, which averages out in case the grid has various wind directions throughout the time domain. Strong enhancements in the coverage map therefore indicate artefacts.

The result of the Gaussian peak-fitting algorithm can now be compared to CO inventories and other methods. For the Cairo location, EDGAR and DACCIWA are used. Furthermore, results from 2019 from the CSF and IME methods are used (personal communication with G. Leguijt). The results are shown in Figure 5.30. The divergence method estimate (520 [420-620] Gg/a) is similar to the DACCIWA estimate (550 Gg/a). The divergence estimate deviates 20% from both the CSF and the IME method. The EDGAR quantification seems to be strongly underestimated.

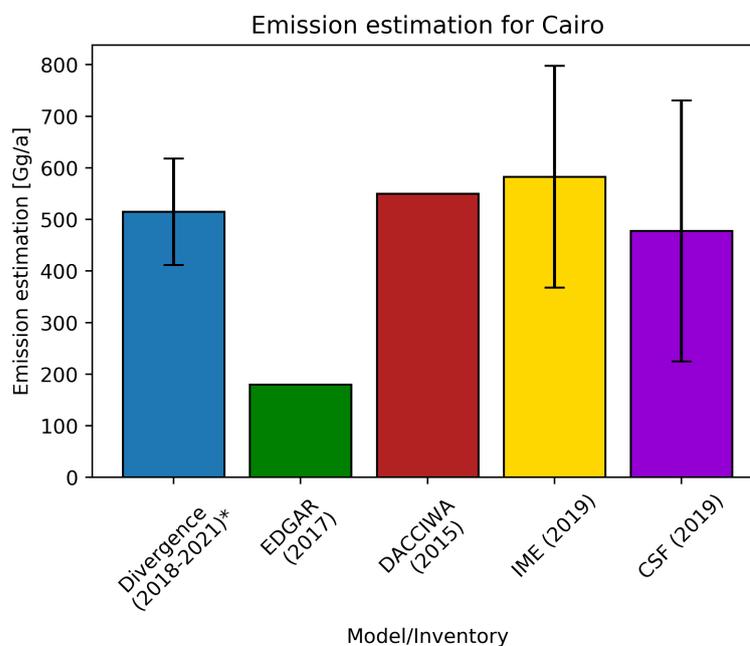


Figure 5.30: Emission quantification estimates for the divergence method, compared to different inventories and quantification methods. A strong underestimate is seen for EDGAR, while DACCIWA and IME/CSF are within the uncertainty of the divergence estimate. * Results for the divergence method are shown for 01/July/2018 - 30/June/2021.

Concluding, the Cairo case study has multiple expected sources within the city of Cairo, which are not separated by the divergence method. The city and the sources are identified as one source. The quantification of the emissions is in line with estimates from other methods (CSF/IME) and bottom-up inventories. The divergence output showed an enhancement with partly Gaussian behaviour which was not identified as a source by the Gaussian peak-fitting algorithm. This case study has shown that the enhancement was related to an artefact, and therefore the algorithm successfully ignored the enhancement. This concluded that the algorithm is successful in identifying artefacts.

5.2.5. Riyadh, Saudi-Arabia

The Riyadh domain is taken as $1.2^\circ \times 1.2^\circ$ around the Riyadh city center. The city is located at 500 meters of altitude, while the grid sees elevation reaching up to 1,000 meters (Figure 5.31). The city is located in a desert area, which indicates low cloud coverage and therefore high TROPOMI coverage year-round. The domain has no rivers or water bodies. With over 7.5 million inhabitants, the city is expected to have significant diffuse urban emissions.

Since the Riyadh area is also taken as case study for the NO_x model (section 1.2), it is possible that the identified sources from the NO_x model also are observable in the CO model. The NO_x model identified four power plants and only one cement plant. It is unlikely that power plants are visible in the CO model, since CO is only emitted during incomplete combustion, which should not be the

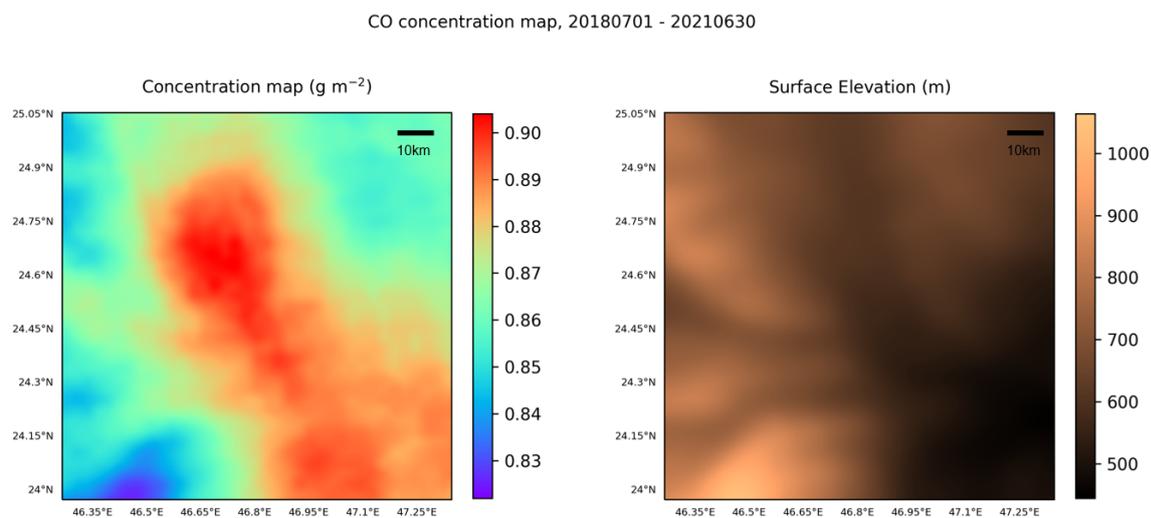


Figure 5.31: The Riyadh area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Large CO concentration enhancements in the center of the grid can be seen.

case in power plants due to the limited incomplete combustion in these plants. The cement plant in the Riyadh area could be observable. In Figure 5.32, the divergence is shown, including the known power and cement plants.

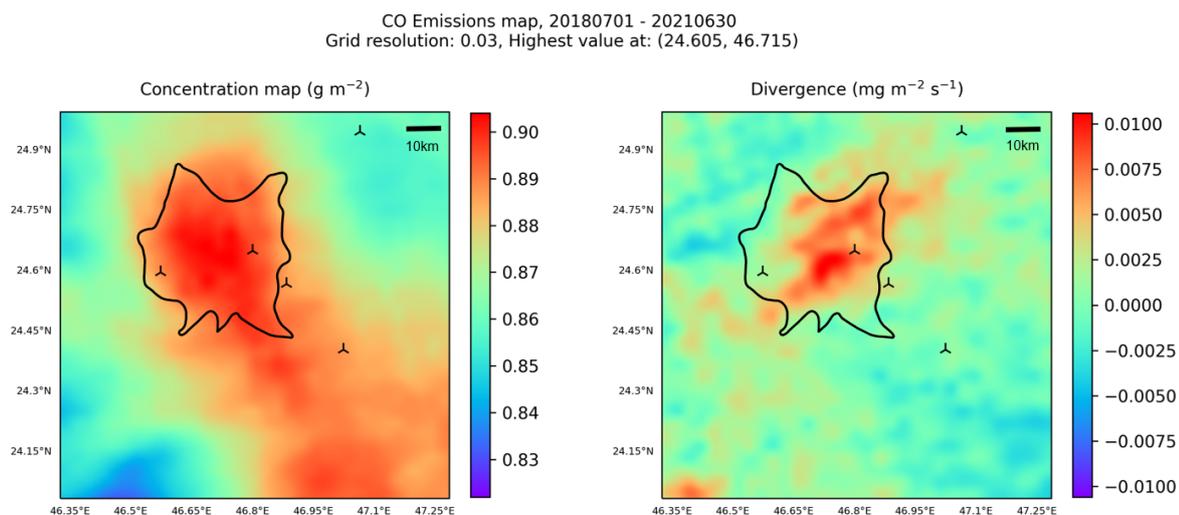


Figure 5.32: The city of Riyadh (outline), showing the TROPOMI CO concentration (left) and the divergence (right) at high resolution (0.03°). The markers indicate emission sources identified by the NO_x model.

None of the known emitters can be matched with enhancements in the divergence map due to the fact that four out of five emitters represent power plants, which have very limited incomplete combustion processes. The most centered marker represents the cement plant. However, this plant is located in close vicinity of the city center. The area around the marker does show some divergence enhancements, which is most likely due to the excessive background signal from the city of Riyadh. Therefore, it cannot be directly related to the cement plant.

When looking at the coverage maps (Figure C.31 and Figure C.33), no artefact indications are observed. The coverage map does show some enhancements towards the south of the grid, but these are not directly related to enhancements in the concentration map. These enhancements can be

classified as noise. Since no individual sources are identified, the Riyadh area is analysed as one (urban) source. The emission quantification is again analysed by using the Gaussian peak-fitting algorithm. The results are shown in Figure 5.33. The location of expected emissions from EDGAR nicely matches with the results from the Gaussian peak-fitting algorithm. The estimate from the divergence method is slightly higher than the EDGAR estimate, but within the uncertainty of the model: 250 [200-300] Gg/a versus 200 Gg/a.

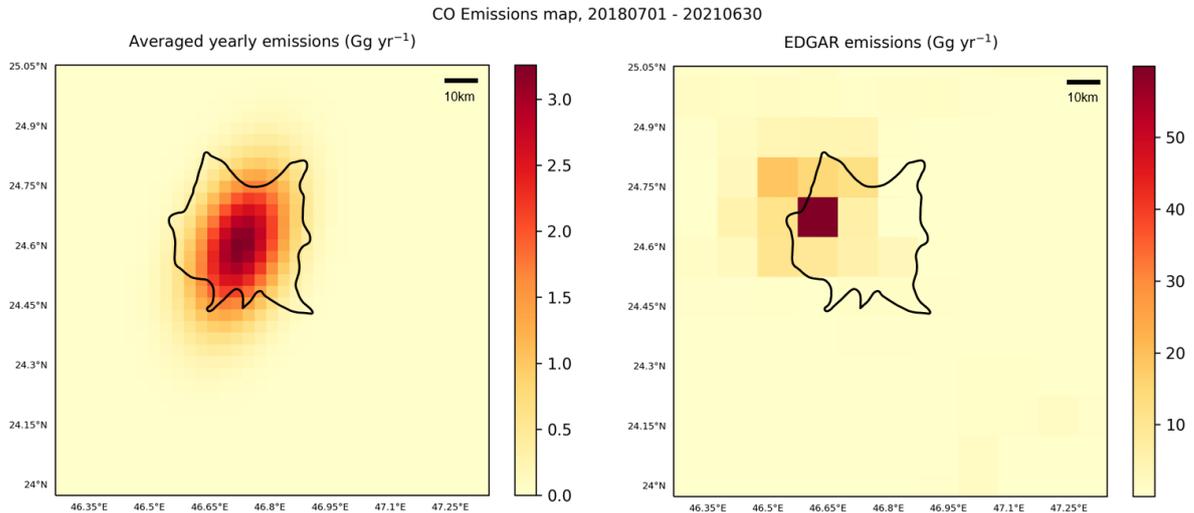


Figure 5.33: The resulting emissions estimate for the divergence method (left) and the EDGAR inventory estimate (right). The divergence output perfectly matches the city outline, while the spatial allocation of emissions for EDGAR shows a slight mismatch.

The result can be compared to other methods and inventories (Figure 5.34). Both the CSF and IME quantification estimates fall within the uncertainty of the divergence estimate.

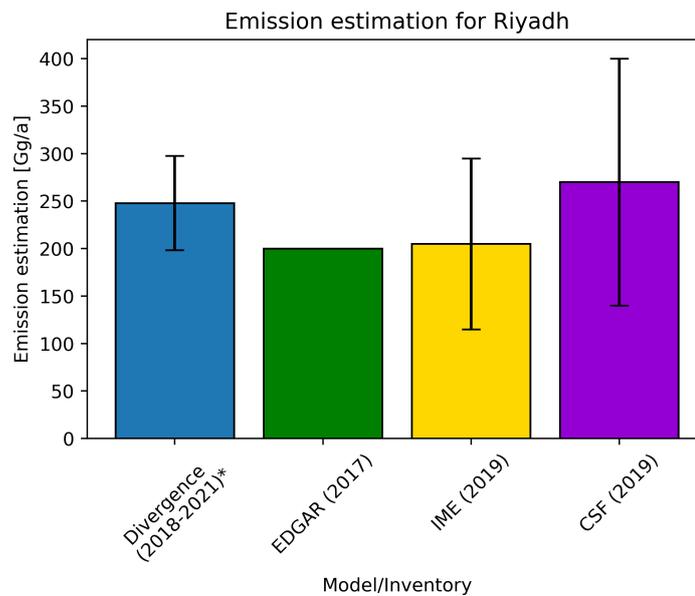


Figure 5.34: Emission quantification estimates for the divergence method, compared to the EDGAR emission inventory and results from the IME/CSF quantification methods. All emission estimates are within the uncertainty of the divergence estimate. * Results for the divergence method are shown for 01/July/2018 - 30/June/2021.

Lastly, the daily coverage is used to determine whether the Riyadh location shows any seasonality in terms of grid coverage. As can be seen in [Figure C.33](#) in [Appendix C.6](#), Riyadh has near-perfect coverage in the summer months. During the winter months, worse coverage is observed. However, this degradation in grid coverage is minimal. Overall, the Riyadh location has good coverage year-round.

Since Riyadh has good coverage year-round with very little seasonal variability, this case study can be used to observe the seasonality of the CO concentration and CO emissions in the grid for the year 2020. Since the divergence method needs more than one day of data for a sufficient analysis, an interval of 90 days is used as a running window. For the analysis of March 1st, the time domain [January 16th, April 15th] is used. To analyse every day in the year separately, 366 model runs with a total of 456 days are used (2020 is a leap year) to calculate the mean CO concentration of the time domain, where every model run uses 90 days of data. Furthermore, the Gaussian peak-fitting algorithm is used to estimate the emissions from the time domain. Using only 90 days of data may be insufficient for the algorithm to reach a stable output. However, since Riyadh has good coverage year-round, it is expected that each gridpoint in the grid reaches the coverage threshold of 20 data pixels for 90 days of data. The day-to-day analysis is used to explore the limits of the divergence model for CO, and the output of the analysis is not expected to be fully reliable for all days.

The results for the 2020 analysis are shown in [Figure 5.35](#). Two things can be concluded from the figure. First, the background-corrected CO concentration has a slight variation throughout 2020. The highest concentration is observed in the spring, with a yearly low during the summer months, which is consistent with findings from research by Té et al. [76]. This research found that CO surface concentration has a maximum around February and a minimum around July. The maximum concentration in February can mainly be explained by the Planetary Boundary Layer (PBL) height being at its maximum and OH oxidation being at its weakest. For July, the opposite is found. Second, the emission estimate by the divergence method clearly shows a large dip in the emission estimate for the city of Riyadh for the months February - March. The figure shows an anti-correlation between emissions and concentration in 2020. Since CO has a lifetime of 3 months ([chapter 2](#)), the summertime emissions can lead to the increase in CO concentration up to October.

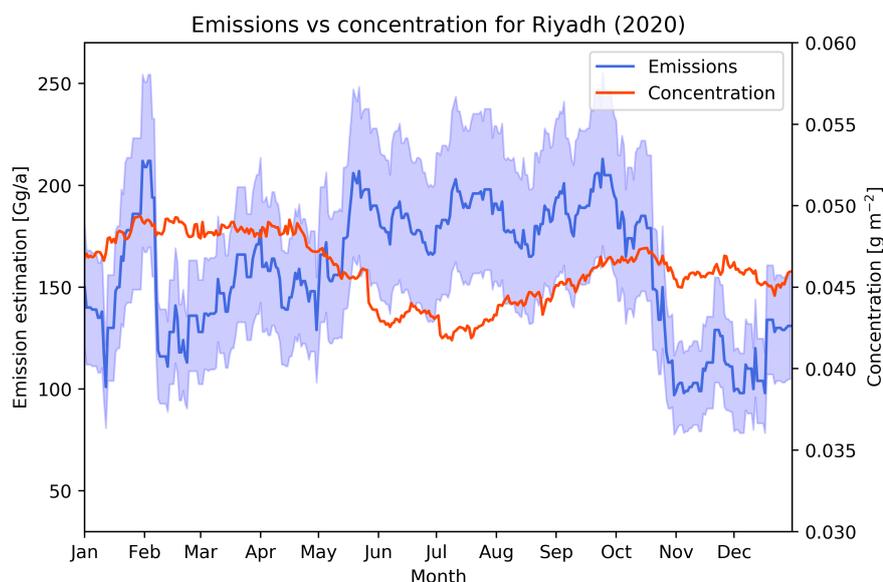


Figure 5.35: Mean background-corrected CO concentration (red) and estimated CO emissions (blue) for Riyadh in 2020 using a 90-day running mean. The blue band shows the uncertainty for the emission estimate. The strong dip around February might be due to coverage issues.

The dip in emissions can have multiple explanations. First of all, the Government of Saudi-Arabia issued a nation-wide lockdown in March 2020, which could have resulted in lower CO emissions and concentrations in that time period [77]. Due to this lockdown, it is expected that emissions from the city of Riyadh decrease after March 2020. Due to the lifetime of CO of 1-3 months, it is expected that the CO concentration in the domain decreases between April-June, following a lockdown starting in March. From the figure, it can be seen that the CO emission estimate dips early March, while the CO concentration dips after spring, which might be lockdown-induced.

Second, city of Riyadh consists of multiple sources, ranging from individual cars to cement plants in the vicinity of the city. However, the Gaussian peak-fitting algorithm identifies the city as one source. Therefore, the city often does not show a plume (like the sources in Jharkhand), but has a steady urban CO concentration due to the emissions from multiple sources. Due to this steady enhancement over the city of Riyadh, the divergence quantity is highly dependent on the wind speed. Strong winds at point-sources result in the emissions being dissipated as a plume. A stronger wind will result in a plume with a lower concentration, which averages out in the divergence calculation. However, in case a source has a steady concentration, the wind speed has a strong influence on the divergence around the source.

Third, the emission estimate by the peak-fitting algorithm can be influenced by low coverage. To quantify the coverage, the average pixels per gridcell are calculated. A day with perfect coverage will result in every gridcell having one pixel as coverage.

To explain the varying emission estimates, the grid coverage and wind speeds for the Riyadh location are observed, as shown in Figure 5.36. The overall grid coverage is good, with the lowest coverage in the winter months. The lowest coverage is seen to be well above the coverage limit of 20 pixels. The summer months experience the best coverage. Next, the wind speed shows minor variations, with a maximum around April and a minimum around September. However, both the minimum and maximum are within 1 m/s from the mean. Therefore, it is unlikely that the large variations in emission quantification are mainly due to the varying mean wind speeds. The lower coverage around the non-summer months might influence the performance of the Gaussian fit. Since higher coverage is assumed to be better, the estimation in the summer months can be taken as more reliable. The dip shown in Figure 5.35 can now be explained due to the lower grid coverage in the winter months. Since the yearly analysis uses a running mean with a window of 90 days, the difference in coverage between January - March and March - May might result in unreliable results.

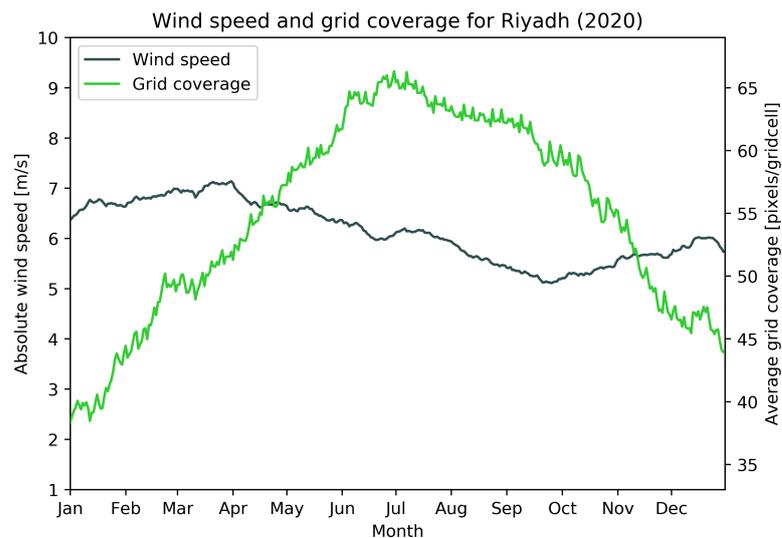


Figure 5.36: The daily mean wind speed (gray) and grid coverage (green) for the domain of the Riyadh case study in 2020. A strong seasonality for grid coverage can be seen. The wind speed varies slightly.

To test the validity of the results of 2020 and to see if there are any Covid-19 related issues, the data for 2019 is studied as well. The results are shown in [Figure C.34](#) and [Figure C.35](#) in [Appendix C.6](#). The emission plot shows a similar anti-correlation between emissions and concentration for the Riyadh area in 2019. Furthermore, the wind speed and grid coverage have a similar pattern as the 2020 data.

To better observe the differences between the two years, a relative emission and concentration plot is used ([Figure 5.37](#)). The figure shows that CO concentrations in 2020 were higher than in 2019 up until March (2020). From early April, CO concentrations in 2019 were higher. This might be due to the effect of the Covid-19 related lockdown. This imbalance shifted in early fall, when 2020 saw higher CO concentrations. The emissions in 2020 were higher than 2019 up until early August. From early fall, the observed emissions are closely similar, with the exception of early winter, where 2019 emissions were higher. The divergence method calculated an emission surplus of 29 Gg for 2020 as compared to 2019.

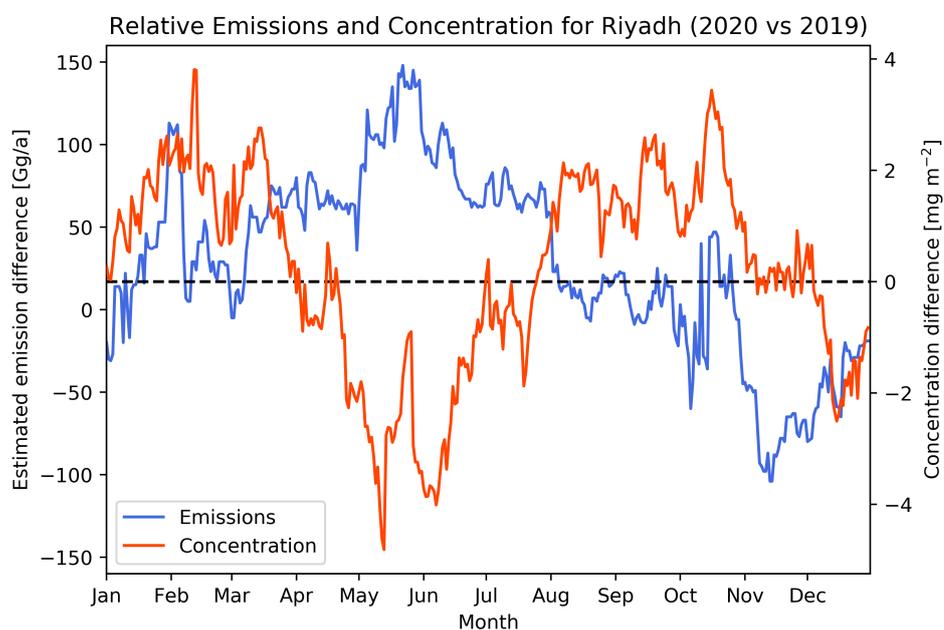


Figure 5.37: The emission (blue) and concentration (red) difference between 2020-2019 using a 90-day running mean. A positive value means larger emissions and higher concentrations in 2020. A dip in relative concentrations can be clearly seen after March 2020.

Concluding, the Riyadh case study has multiple known sources in the grid, as taken from the study by Beirle et al. [1]. However, since most of these sources are power plants, it is expected that these sources do not turn up in the divergence map. As can be seen in [Figure 5.32](#), none of the marked locations are identified by divergence method. The Gaussian peak-fitting algorithm identifies one source, which is the city of Riyadh. The emission quantification from the divergence method is in line with other methods (IME/CSF) and the EDGAR inventory. The divergence method has good performance in terms of emission quantification for the Riyadh case study. Furthermore, the Riyadh case study showed that 90 days of data with good coverage is sufficient for the divergence method to have a stable output for the Gaussian peak-fitting algorithm. The case study showed that emission changes due to a Covid-19 lockdown are possible to detect with the divergence model.

5.2.6. The Influence of the Covid-19 Pandemic on CO Emissions in Wuhan

To investigate the sensitivity of the divergence method to changing CO emissions, Wuhan is studied. This case study will mainly focus on the emissions estimated from the divergence model for 2019 and 2020. As the city of Wuhan went in lockdown early 2020, the steel plants in the area were not closed down. However, due to a decrease in demand for the remainder of 2020, the production was scaled down [78]. Therefore, it is expected that the divergence method will detect CO emissions from steel plants in the Wuhan area for both 2019 and 2020, albeit with a lower estimate for 2020.

The CO concentration and divergence maps (Figure 5.38) show three possible emissions sources, which match the locations of known steel plants in the area. The plants are identified as the Qinshan plant, the Echeng plant and the Huangshi plant. All three plants belong to the Wuhan Iron and Steel Corporation (WISCO), which was ranked 11th in the 2015 world ranking by volume [79].

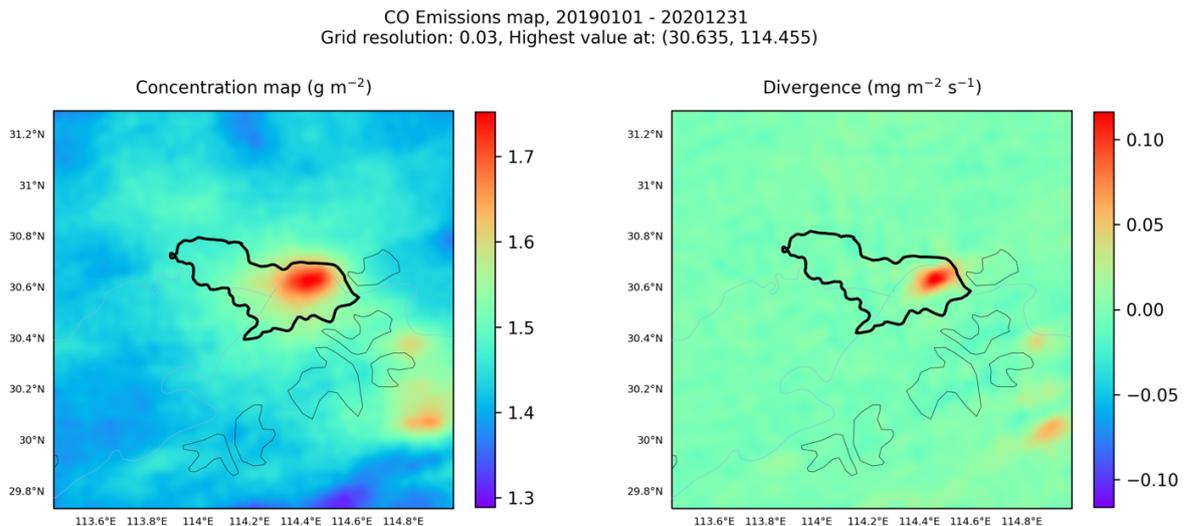


Figure 5.38: The Wuhan area, showing the concentration map (left) and the corresponding divergence map (right) with the contours of the city of Wuhan (centered) and smaller lakes. Three individual sources can be seen.

The results from the Gaussian peak-fitting algorithm can be compared to the REAS inventory, for which the emission estimates are shown in *Industry* and *Total* (Figure 5.39). According to REAS, the majority of the CO emissions in the Wuhan region are due to industry. The divergence method results in a lower estimate than the REAS inventory, unlike the Jharkhand case where the divergence method and REAS were consistent.

The REAS inventory is using estimates from the year 2015. The Chinese government has been reducing steel production nation-wide since 2016 [80]. Since the divergence method uses data from 2019 and 2020, it is likely that the plants emitted less CO due to limited steel production as compared to the year 2015. Furthermore, WISCO cut production of iron and steel by a total of 4.42 million metric tonnes after heavy net losses in 2015. The lower emission estimates from the divergence model as compared to the REAS inventory can be explained due to this production cut [81].

Furthermore, Figure 5.39 shows that the emission estimates for 2019 are higher for the Qingshan and Huangshi plants as compared to the year 2020. However, the quantifications are within the uncertainty bounds for all estimates. The mean emissions for 2020 are 13% lower than the mean emissions for 2019. The emission estimate for the Qinshan plant is 21% lower than the estimate for 2019. Since this is on the outer edge of the uncertainty interval, it is likely that the plant actu-

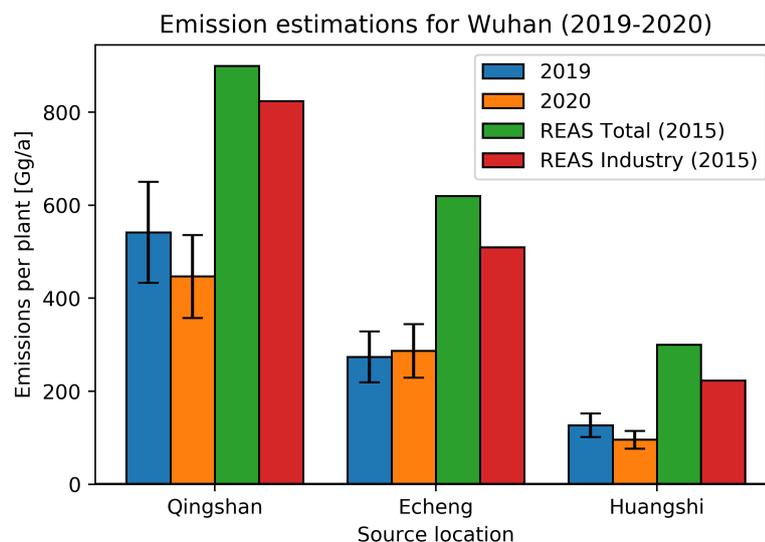


Figure 5.39: Emission quantification estimates for different sources in the Wuhan area for 2019 and 2020 (divergence method) and 2015 (REAS). The REAS estimates are divided into *Industry* and *Total*. The divergence estimates are significantly lower than the REAS estimates.

ally emitted less CO in 2020. The divergence method results confirm that the plants have not been closed in the year 2020. Also, it is likely that the plants produced less steel in 2020 due to a decrease in demand due to the ongoing Covid-19 pandemic.

Concluding, the divergence method successfully confirmed that the steel plants in the Wuhan area were not, or only partly, shut down during the lockdown in 2020. It showed a stable output for only one year of data, which confirms that yearly emission analyses for point-sources are feasible.

5.2.7. Testing the Methane Model for Korpezhe, Turkmenistan

To verify whether the divergence model is functional for methane, a domain in Turkmenistan is analysed. Since the methane model is not optimized and tuned, the model uses settings from the CO model: ERA5 100 m for wind input, a background correction, no altitude correction and an AOT filter $\tau < 0.1$ [82]. The location is chosen since the Korpezhe oil field is extensively studied. The case study is used to test whether the methane model is functional and is able to identify methane sources without being optimized. The locations of identified sources can then be used as input for high-resolution instruments, such as Hyperspectral PRecursor of the Application Mission (PRISMA). PRISMA has a higher resolution than TROPOMI (30×30 m) but a single scene size of 30×30 km, and therefore can be guided by TROPOMI observations to select a proper scene.

The domain for the analysis is chosen as $1.5^\circ \times 1.5^\circ$, centered at (38.1, 54.0). In Figure 5.40, the divergence model output is shown. In the concentration map (left), the high background value of methane due to the long lifetime in the atmosphere can be clearly seen. When looking at the CO case studies, concentration enhancements of 1.5 g m^{-2} with a background of 1 g m^{-2} were often seen. For methane, similar concentration enhancements are seen (0.5 g m^{-2}), but with a background of 10 g m^{-2} . This lower relative enhancement (50% for CO versus 5% for methane) makes methane more challenging for a divergence analysis, as compared to CO.

The Gaussian peak-fitting algorithm identifies three possible sources in the divergence map (Figure 5.40). However, since the algorithm was not able to successfully fit a Gaussian to the data, it

is not known whether the divergence enhancements can be identified as emission sources. This can be explained due to the lack of Gaussian behavior of the divergence enhancements. When using previous research about emission sources in the Korpezhe region, the sources can be identified. The enhancements in the divergence map all match known emission sources from research by Irakulis-Loitxate et al. [82]. The locations, as identified by the divergence method, are shown in Table 5.8.

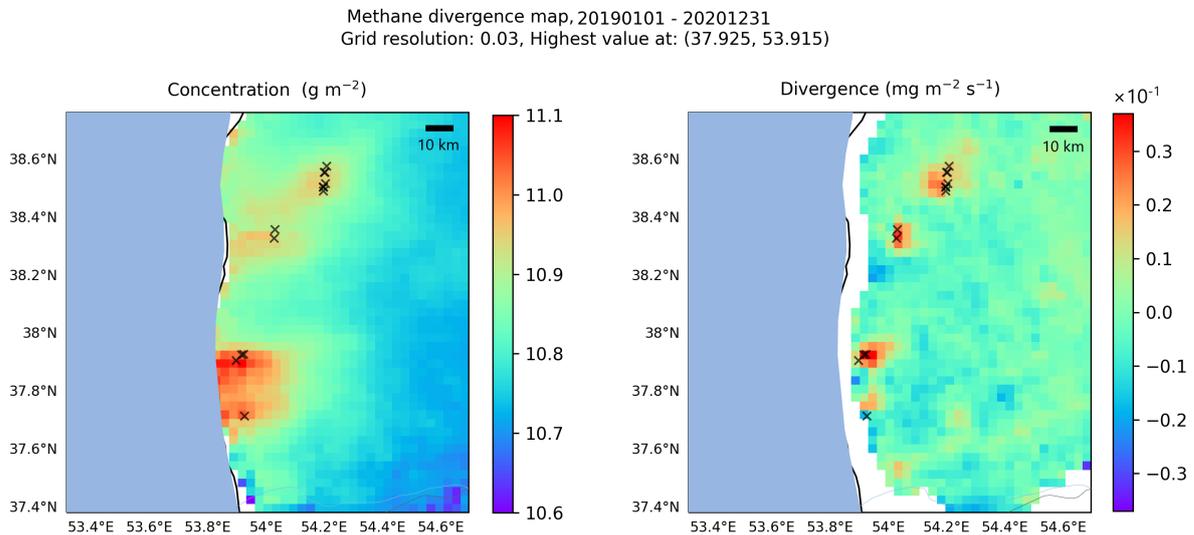


Figure 5.40: The Korpezhe region (Turkmenistan-West), showing methane concentration (left) and the corresponding divergence (right) for 2019 and 2020. Since no sources are quantified by the Gaussian peak-fitting algorithm, known locations of emission sources are shown by ‘x’ [82].

Table 5.8: Three methane emission sources as localized by the divergence method (lat, lon). Since the model was not able to successfully fit a Gaussian to the data, no emission estimates are included. The sources are identified using research from Irakulis-Loitxate et al. [82].

Field	Latitude	Longitude	Source type	Number of sources
Korpeje	38.48	54.19	Ground flare	7
Gamyshlja Gunorta	38.37	54.06	Pit flare	2
Keymir	37.93	53.92	Ground flare/pipeline	3

To verify whether the divergence method can be used to guide high-resolution remote-sensing instruments, PRISMA observations are used around one identified source location (38.48, 54.19). A clear plume is detected in 2020 (Figure 5.41), being emitted from a methane pipeline in the Korpezhe region (personal communication with P. Bijl). The output is plume-masked, which results in only the plume being visible. The methane background is filtered out. This result shows the added value of the divergence method in combination with high-resolution instruments for methane. Furthermore, it can be concluded that the methane model is functional in terms of identifying divergence enhancements, but the Gaussian peak-fitting algorithm is not able to detect all emission sources based on these enhancements. The output can be improved by optimizing the methane model in a similar fashion as the CO model.

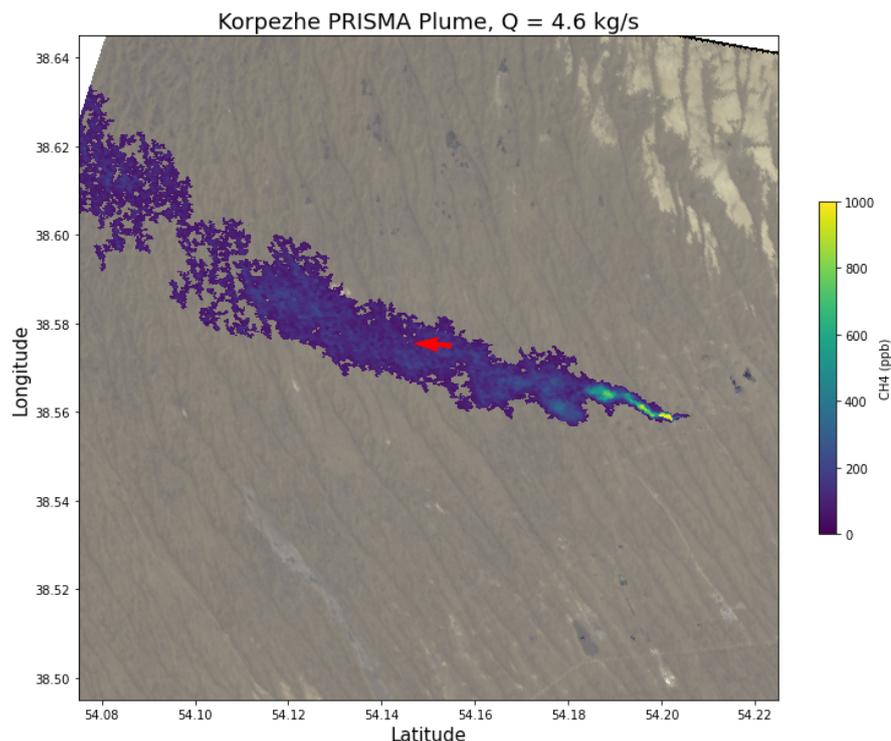


Figure 5.41: A PRISMA methane observation on July 21st, 2020, showing a clear plume from a ground flare at (38.56, 54.20). The red arrow denotes the GEOS-FP 10 m wind direction (personal communication with P. Bijl).

5.2.8. Conclusions from Case Studies

The case studies are used to test the performance of the divergence method with TROPOMI data on various locations around the world, and can therefore be used to answer the research question (section 3.2). In Table 5.9, the main findings from each case study are shown, which are used in chapter 6 to conclude whether the research question has been answered.

Table 5.9: The case studies with demonstrated performance conclusions.

Name	Country	Conclusions from case studies
Duisburg	Germany	Source separation up to 13 km in distance demonstrated, quantification results in accordance with inventories.
Hebei	China	Shown the value of the divergence method in regions like China, where limited information about sources is available.
Jharkhand	India	Performance of altitude correction demonstrated, quantification results in accordance with inventories.
Cairo	Egypt	Successful artefact identification, cities can be identified as a single source.
Riyadh	Saudi-Arabia	Ability to perform a day-to-day analysis demonstrated, all inventory/quantification estimates within uncertainty of model.
Wuhan	China	Ability to compare yearly emissions, lower CO emissions in Wuhan in 2019/2020 compared to 2015.
Korpezhe	Turkmenistan	Divergence model for methane is functional, divergence method can be used to guide high resolution instruments.

6

Conclusion

The main research question of this thesis, as introduced in [section 3.2](#), is:

"How can the CTM-independent emission quantification and localization method for NO₂ by Beirle et al. [1] be adapted for carbon monoxide so it can be used for identification of closely-spaced emission sources and how can this model be optimized and utilized?"

It can be concluded that the research question for this research has been answered. A fully functioning model for CO and methane has been developed, where the CO model is fully optimized using pseudo data. It is concluded that the model is able to successfully identify and separate closely-spaced emission sources. The sub-questions are answered below.

The three trace gases (NO_x, CO and methane) are different in mixing ratio and lifetime. The short lifetime for NO_x mainly resulted in a low global background mixing ratio (compared to CO and methane) with strong enhancements over emission sources. The corresponding divergence of these enhancements is found to be large, which results in a strong output in the divergence map. This strong signal makes it fairly easy for the Gaussian peak-fitting algorithm to detect sources. The lower enhancements for CO make that the CO divergence model experiences more trouble in identifying the source. However, even though the divergence map shows a weaker output, the Gaussian peak-fitting algorithm is still able to identify emission sources. Furthermore, the higher background for CO results in the CO model needing a background and altitude correction. Lastly, the CO model uses a land-mask filter, since CO measurements above waterbodies are hard to interpret.

The combination of large enhancements for the NO_x model and uncertainties in the wind direction of the ERA5 data make that the NO_x model's main error source is stripes in the divergence. This decreases the signal-to-noise ratio of the divergence output and leads to erroneous divergence values. Furthermore, it leads to an incorrect emission estimate. The case studies showed that the main error source for the CO model is the inaccuracies in altitude correction. These inaccuracies lead to artefacts in the divergence output. Lastly, the CO model showed noise due to stripes in the divergence. However, this is found not to be as significant as for the NO_x model, due to the lower enhancements of CO.

The method by Beirle et al. [1] has two main error sources: (1) inaccuracies due to the numerical differentiation method, and (2) inaccuracies due to the Gaussian peak-fitting algorithm. The algorithm that Beirle uses results in errors due to the many ways to fit a curve to data. The algorithm uses bounds for width and center to ensure a peak-fitting procedure. However, this decreases the

accuracy of the fit since it often leads to an overestimation of the data. Since the derivative of the flux can only be numerically estimated, an error for the flux calculation cannot be prevented.

Uncertainties in the meteo data consist of two parts: uncertainties in the wind speed and uncertainties in the wind direction. If the direction of a plume does not match the direction of the wind, this results in a stripe-like pattern in the divergence output. This erroneous stripe pattern influences the signal-to-noise ratio of the divergence, which decreases the performance of the Gaussian peak-fitting algorithm. Uncertainties in the wind speed result in erroneous emission quantifications. The meteo data is especially uncertain below 1 m/s wind speed. Therefore, these values are filtered out by the model. Wind speeds above 10 m/s lead to broken-up plumes, which causes dipoles in the divergence. Therefore, these values are also filtered out.

The performance of the model is tested by using pseudo data, generated by the WRF-chem CTM. The emission estimate from the divergence method using the pseudo data as input can be compared to the actual emissions that were used in generating the pseudo data. A sensitivity analysis is used to test the performance of the model for different filter settings, after which the optimal settings can be selected. This resulted in an optimal model. The sensitivity analysis resulted in six conclusions. (1) A filter on aerosol optical thickness $AOT < 0.5$ must be used. (2) A filter on height scattering layer $HSL < 5$ km must be used. (3) The optimal model resolution for emission quantification is 0.04° . (4) The optimal model resolution for source separation is 0.02° . (5) The optimal wind field is ERA5 100 m with a wind speed filter on < 1 and > 10 m/s, and (6) a filter on TROPOMI resolution is not included, since it does not improve the performance of the model.

The model resolution has influence on two main aspects of this research: the ability to separate closely-space emission sources, and the quantification of these sources. A sensitivity analysis is used to test the performance of different model resolutions on the emission quantification. It is concluded that the optimal resolution for emission quantification is 0.04° , which is higher than the TROPOMI resolution. Second, a test for the ability to separate closely-spaced emission sources resulted in the conclusion that the source separation performance is directly related to the model resolution. The model performs best for the highest possible model resolution, which was found to be 0.02° . Therefore, the optimal resolution for both source separation and emission quantification is found to be 0.03° .

The pseudo data analysis concluded that worse coverage results in a worse performance for the divergence method. The near-perfect coverage for NO_x makes that the divergence model has a stable output for a limited number of data days. Influence of coverage is mainly related to the number of data days that are needed for the divergence to obtain a stable output. For the CO model, the lower coverage (compared to NO_x) resulted in the model needing at least 20 individual days with high coverage to obtain a stable signal in the divergence map. The case studies showed that the coverage of CO has a large seasonal dependence. This makes that quarterly (or monthly) analyses of the divergence data are infeasible, simply because the coverage for such a limited time interval is insufficient. The Riyadh case study showed that a period of 90 days is sufficient. For the analysis to have a stable output, however, this 90-day period does need to include at least 20 high coverage days, which is the case for the Riyadh study. From this, it can be concluded that for worse-coverage locations, even more days need to be used for a satisfying result.

To estimate the uncertainty of the model, the pseudo data with the optimal model parameters is used. This test showed that the model has a slight overestimation for the emission quantification, with a standard deviation of 20%. Therefore, the deviation of 20% is used as uncertainty for the model.

Two numerical methods are tested. The fourth-order central finite difference methods results in lower grid coverage, but is more accurate. The second-order central finite difference method results in better grid coverage, but with lower accuracy. It is concluded that the fourth-order method is preferable due to the better accuracy. However, for low-coverage domains, the second-order method is preferable. A domain is said to be low-coverage if the Gaussian peak-fitting algorithm does not result in a successful fit due to coverage issues.

From the case studies, it is concluded that the Gaussian peak-fitting algorithm performed well in identifying individual emission sources. The pseudo data analysis showed that the quantification limit for accurate emission estimations is 50 Gg/yr, which makes it able to detect a large number of emission sources world-wide. The algorithm does, however, detect artefacts in rare cases. Therefore, it is necessary to check the geographical locations of identified sources with emission inventories and/or satellite imagery. Furthermore, the case studies demonstrated that the Gaussian peak-fitting algorithm can separate sources up to 13 km in distance. When translating this to the TROPOMI pixel resolution, this would be equal to approximately 2 pixels in distance. Therefore, it is concluded that the divergence method, in combination with the Gaussian peak-fitting algorithm, performs very well. Lastly, the divergence method successfully identified five large emission sources in the Hebei region without any a priori information on possible source locations, showing the added value of the divergence method in regions where limited information on emission source locations and quantifications is available.

This thesis successfully developed and optimized a divergence model for carbon monoxide. The use of this model, in combination with TROPOMI atmospheric trace gas measurements, can be an effective tool in climate change mitigation. With the rising global mean temperature due to increasing greenhouse gas and air pollutant emissions, it gets increasingly important to accurately quantify and localize emission sources, as the significant reduction of these emissions can be the decisive action in slowing and ultimately reversing the process of global warming.

7

Recommendations

From the report, it is concluded that the research objective has been met. However, the research showed imperfections in the model and the results showed promising future applications for the divergence method. This chapter describes the recommendations for future research and the recommendations for possible applications, which have not been included in this research.

7.1. Recommendations for Future Research

The divergence method uses an altitude correction to correct for the correlation between CO concentration and surface elevation. This correction procedure uses the CO concentration at different surface altitudes for individual days, after which linear regression is used to find a correction factor. To include seasonality in this factor, a sine function is fitted to the individual daily correction factors. As the case studies showed, this correction improves the divergence output. However, as the Jharkhand case study showed, the correction does not fully solve the problem. The domain showed divergence enhancements that could be related to altitude differences. The seasonality plots of the altitude correction factor showed large variations between the daily altitude correction factors. Therefore, it is recommended to improve the linear regression procedure, to decrease the variation in daily correction factors. This will also improve the seasonality fit, which increases the accuracy of the altitude correction.

Currently, the method requires that every new grid location first needs a separate model run to determine the altitude correction factors. This is due to the fact that there is no global altitude correction map for CO. As this has previously been done by Chen [83], it is recommended to design a similar map for CO. This will decrease the workload for the divergence model and will allow the possibility to make a global CO emission map by using the divergence method.

The divergence method uses a Gaussian for the peak-fitting algorithm. This research concluded that the Gaussian fit is robust for the identification of emission sources. However, this research did show an overfit when using the Gaussian for many case study locations. Therefore, it is recommended to research other methods. Two suitable options for peak fitting are the Lorentzian and the Voigt function. Since the Gaussian has a limit in separating point-sources due to the width of the Gaussian, the Lorentzian function might also be better performing in separating closely-spaced point-sources, due to the steeper slope of the Lorentzian function.

The divergence method currently researched two numerical differentiation methods: the second-order and fourth-order central finite difference. However, no combination of numerical methods is investigated. To improve grid coverage, a combination of forward and backward Euler methods can be used, in combination with the central finite difference methods. This will be feasible in low-coverage locations, where grid coverage outweighs model accuracy.

This research used three different wind fields for the analysis: GEOS-FP 10 m, ERA5 10 m and ERA5 100 m (altitude) winds. However, it is recommended for future research to investigate the performance of the model for different wind altitudes, especially the 300 m and 450 m winds. Since the divergence method mainly is concerned about the initial start location of a plume, the 10 m and 100 m are viable options. However, plumes can reach higher than 100 m in the atmosphere. Therefore, it must be investigated how different wind altitudes relate to noise in the grid, as well as how the different wind fields perform in source identification.

The CO model is tuned by using WRF-generated pseudo data. However, only one domain included closely-spaced point sources. The WRF pseudo data resulted in only nine feasible analysis locations. Therefore, it is recommended to tune the divergence model with a larger number of pseudo data locations, as well as using locations that vary in coverage, source strength, orography and climate. This makes the pseudo data analysis more robust and the conclusions from the analysis more reliable.

The CO model is fully tested and optimized, after which it is used to research six case studies. The methane model, however, is fully functional but not optimized. The functionality is tested by analysing a domain in Turkmenistan. It is recommended that future research will investigate the need for albedo and aerosol corrections (similar to the CO altitude correction) for the methane divergence model. Furthermore, the methane model needs to be extensively tested and optimized in a similar manner as the CO model. This can be done by using pseudo data and various sensitivity analyses to test the model performance for different model settings.

7.2. Recommendations for Applications

Since the divergence method is able to identify, quantify and separate point-sources, it can be used to build an inventory of emission sources. With TROPOMI's daily global coverage, the divergence method can be used to establish a world-wide inventory of emission sources. Since the divergence method has a run time of approximately 23 minutes for a year of CO data using a regular 40×40 grid on a 4-core computer, many domains can rapidly be analysed. When using a supercomputer, like Snellius (4,608 cores), a similar grid can be analysed in under a second. This opens up possibilities of rapidly analysing world-wide grids. As the case studies pointed out, the locations of estimated emissions in the inventories (especially EDGAR and REAS in China) are not always correct. Furthermore, many sources in China were not included in the databases. Therefore, the divergence method can help identifying missing sources, and can be used to obtain an order of magnitude of emissions of these sources. The divergence method can be used to tune the bottom-up inventories, as is done in this research for the case studies. Since the divergence method can use a large variation of grid sizes, it can be used to analyse the full globe for CO emissions. Therefore, it is recommended that future research investigates the ability of the divergence method to run large grids, for instance continents, in one model run.

As mentioned in the introduction, the rising temperatures on Earth make it increasingly important to accurately monitor and control the emission sources of greenhouse gases and other pollutants. The divergence method can improve the process of monitoring global emissions. Therefore, the divergence method can play an important role in climate change mitigation. The method can be combined with high-resolution instruments, for instance the PRISMA instrument. The divergence method is able to identify large grids of emissions, which can be used to guide the high resolution instruments, which have a higher spatial resolution (up to 30m^2) but a smaller spatial domain (around 30km^2). This low-high resolution instrument combination has previously been researched by Varon et al. [84] for methane. This research recommends to investigate the possibility of combining the TROPOMI divergence method with high resolution CO instruments.

Data Availability

TROPOMI CO data are available at <https://s5phub.copernicus.eu/dhus/#/home>.

TROPOMI CH4 data are available at https://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/ch4/14_14_Lorente_et_al_2020_AMTD/.

GEOS-FP wind data can be downloaded at gmao.gsfc.nasa.gov/GMAO_products.

ECMWF ERA5 wind data can be downloaded at <https://cds.climate.copernicus.eu>.

EDGAR v5.0 emissions are available at edgar.jrc.ec.europa.eu/overview.php?v=50_GHG.

REAS v3.2 emissions are available at <https://www.nies.go.jp/REAS/index.html#data%20sets>.

E-PRTR emissions are available at <https://industry.eea.europa.eu/download>.

TNO emissions are available at <https://atmosphere.copernicus.eu/catalogue#/>.

DACCIWA emissions are available at <http://baobab.sedoo.fr/DACCIWA/>.

The developed models and model output are available upon request via email:

L.R.Foorthuis@sron.nl / L.R.Foorthuis@gmail.com (author) or

J.D.Maasakkers@sron.nl (supervisor)

Bibliography

- [1] S. Beirle et al. “Pinpointing nitrogen oxide emissions from space”. In: *Science Advances* 5.11 (2019). DOI: [10.1126/sciadv.aax9800](https://doi.org/10.1126/sciadv.aax9800).
- [2] D. J. Jacob. *Introduction to Atmospheric Chemistry*. Princeton University Press, 1999.
- [3] European Space Agency. *Report For Mission Selection, Carbonsat*. ESA, 2015.
- [4] S. T. Kenea et al. “Comparison of XCH₄ Derived from g-b FTS and GOSAT and Evaluation Using Aircraft In-Situ Observations over TCCON Site”. In: *Asia-Pacific Journal of Atmospheric Sciences* 55 (2019), pp. 415–427. DOI: [10.1007/s13143-019-00105-0](https://doi.org/10.1007/s13143-019-00105-0).
- [5] D. J. Jacob et al. “Satellite observations of atmospheric methane and their value for quantifying methane emissions”. In: *Atmospheric Chemistry and Physics* 16 (2016), pp. 14371–14396. DOI: [10.5194/acp-16-14371-2016](https://doi.org/10.5194/acp-16-14371-2016).
- [6] J. P. Veefkind et al. “TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications”. In: *Remote Sensing of Environment* 120 (2012), pp. 70–83. DOI: [10.1016/j.rse.2011.09.027](https://doi.org/10.1016/j.rse.2011.09.027).
- [7] A. J. Dore et al. “Evaluation of the performance of different atmospheric chemical transport models and inter-comparison of nitrogen and sulphur deposition estimates for the UK”. In: *Atmospheric Environment* 119 (2015), pp. 131–143. DOI: [10.1016/j.atmosenv.2015.08.008](https://doi.org/10.1016/j.atmosenv.2015.08.008).
- [8] M. Nurunnabi. “New insights of NO_x sources from the divergence of the mean NO₂ flux”. In: *Journal of Microbiology, Immunology and Infection* 54.1 (2021), pp. 127–128. DOI: [10.1016/j.jmii.2020.07.023](https://doi.org/10.1016/j.jmii.2020.07.023).
- [9] L. Brunner. “Stratospheric ozone and temperature evolution over the past decades”. PhD thesis. July 2014. DOI: [10.13140/RG.2.2.16933.29922](https://doi.org/10.13140/RG.2.2.16933.29922).
- [10] F. K. Lutgens and E. J. Tarbuck. *The atmosphere*. 9th ed. Prentice Hall, 2000.
- [11] T. S. Ledley et al. “Climate change and greenhouse gases”. In: *Eos, Transactions American Geophysical Union* 80.39 (1999), pp. 453–458. DOI: [10.1029/99eo00325](https://doi.org/10.1029/99eo00325).
- [12] Centers for Disease Control and Prevention. *Nitrogen Oxide and Nitrogen Dioxide*. <https://www.cdc.gov/niosh/docs/2003-154/pdfs/6014-1.pdf>. Last access: 26/04/2021.
- [13] D. J. Jarvis et al. *WHO Guidelines for Indoor Air Quality: Selected Pollutants*. World Health Organization, 2010.
- [14] P. J. Crutzen. “The Role of NO and NO₂ in the Chemistry of the Troposphere and Stratosphere”. In: *Annual Review of Earth and Planetary Sciences* 7.1 (1979), pp. 443–472. DOI: [10.1146/annurev.ea.07.050179.002303](https://doi.org/10.1146/annurev.ea.07.050179.002303).
- [15] L. van der Maas. “Evaluating blast furnace emissions using high resolution TROPOMI CO and NO_x measurements and emission inventory data”. PhD thesis. June 2019.
- [16] G. D. Thurston. “Outdoor Air Pollution: Sources, Atmospheric Transport, and Human Health Effects”. In: *International Encyclopedia of Public Health*. Ed. by Harald Kristian (Kris) Heggenhougen. Oxford: Academic Press, 2008, pp. 700–712. ISBN: 978-0-12-373960-5. DOI: [10.1016/B978-012373960-5.00275-6](https://doi.org/10.1016/B978-012373960-5.00275-6).

- [17] W. F. J. Evans, C. T. McElroy, and I. E. Galbally. "The conversion of N_2O_5 to HNO_3 at high latitudes in winter". In: *Geophysical Research Letters* 12.12 (1985), pp. 825–828. DOI: [10.1029/g1012i012p00825](https://doi.org/10.1029/g1012i012p00825).
- [18] J. A. Geddes et al. "Long-Term Trends Worldwide in Ambient NO_2 Concentrations Inferred from Satellite Observations". In: *Environmental Health Perspectives* 124.3 (2016), pp. 281–289. DOI: [10.1289/ehp.1409567](https://doi.org/10.1289/ehp.1409567).
- [19] Food & Rural Affairs Department for Environment. *Concentrations of Nitrogen Dioxide 2021*. Last access: 03/05/2021. 2021. URL: <https://www.gov.uk/government/statistics/air-quality-statistics/nitrogen-dioxide>.
- [20] S&P Global Market Intelligence. *Diesel's decline may not be such good news for environmentalists, carmakers*. 2020. URL: <https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/diesel-s-decline-may-not-be-such-good-news-for-environmentalists-carmakers-56537596>.
- [21] O. Schneising et al. "A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor". In: *Atmospheric Measurement Techniques* 12.12 (2019), pp. 6771–6802. DOI: [10.5194/amt-12-6771-2019](https://doi.org/10.5194/amt-12-6771-2019).
- [22] E. K. Vakkilainen. "Solid Biofuels and Combustion". In: *Steam Generation from Biomass* (2017), pp. 18–56. DOI: [10.1016/b978-0-12-804389-9.00002-2](https://doi.org/10.1016/b978-0-12-804389-9.00002-2).
- [23] WHO Regional Office Denmark. *Air Quality Guidelines - Second Edition, Carbon Monoxide*. World Health Organisation (WHO), 2000.
- [24] A. Voiland. *Fourteen years of carbon monoxide from MOPITT*. 2015. URL: <https://climate.nasa.gov/news/2291/fourteen-years-of-carbon-monoxide-from-mopitt/>.
- [25] X. Faïn et al. "Northern Hemisphere atmospheric history of carbon monoxide since preindustrial times reconstructed from multiple Greenland ice cores". In: (2021). DOI: [10.5194/cp-2021-28](https://doi.org/10.5194/cp-2021-28).
- [26] R. G. Prinn. "THE CLEANSING CAPACITY OF THE ATMOSPHERE". In: *Annual Review of Environment and Resources* 28.1 (2003), pp. 29–57. DOI: [10.1146/annurev.energy.28.011503.163425](https://doi.org/10.1146/annurev.energy.28.011503.163425).
- [27] Nanoplus. *Carbon Monoxide Detection*. <https://nanoplus.com/en/applications/applications-by-gas/carbon-monoxide-co/>. Last access: 14/04/2021.
- [28] A. Ernst and J. D. Zibrak. "Carbon Monoxide Poisoning". In: *New England Journal of Medicine* 339.22 (1998), pp. 1603–1608. DOI: [10.1056/nejm199811263392206](https://doi.org/10.1056/nejm199811263392206).
- [29] D. Wuebbles. "Atmospheric methane and global change". In: *Earth-Science Reviews* 57.3-4 (2002), pp. 177–210. DOI: [10.1016/s0012-8252\(01\)00062-9](https://doi.org/10.1016/s0012-8252(01)00062-9).
- [30] H. Bürgmann. "Methane Oxidation (Aerobic)". In: *Encyclopedia of Geobiology* (2011), pp. 575–578. DOI: [10.1007/978-1-4020-9212-1_139](https://doi.org/10.1007/978-1-4020-9212-1_139).
- [31] IPCC. *Global Warming Potentials (IPCC Second Assessment Report)*. 2021. URL: <https://unfccc.int/process/transparency-and-reporting/greenhouse-gas-data/greenhouse-gas-data-unfccc/global-warming-potentials>.
- [32] D. A. Vallero. "Air pollution biogeochemistry". In: *Air Pollution Calculations* (2019), pp. 175–206. DOI: [10.1016/b978-0-12-814934-8.00008-9](https://doi.org/10.1016/b978-0-12-814934-8.00008-9).
- [33] H. O. Pörtner et al. "Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change". In: *Cambridge University Press. In Press*. ()

- [34] D. Archer et al. “Atmospheric Lifetime of Fossil Fuel Carbon Dioxide”. In: *Annual Review of Earth and Planetary Sciences* 37.1 (2009), pp. 117–134. DOI: [10.1146/annurev.earth.031208.100206](https://doi.org/10.1146/annurev.earth.031208.100206).
- [35] H. Chen et al. “Determinants influencing seasonal variations of methane emissions from alpine wetlands in Zoige Plateau and their implications”. In: *Journal of Geophysical Research* 113.D12 (2008). DOI: [10.1029/2006jd008072](https://doi.org/10.1029/2006jd008072).
- [36] Global Monitoring Laboratory. *Global CH₄ Monthly Means*. Last access: 24/04/2021. 2021. URL: https://www.esrl.noaa.gov/gmd/ccgg/trends_ch4/.
- [37] S. Javadinejad, S. Eslamian, and K. Ostad-Ali-Askari. “Investigation of monthly and seasonal changes of methane gas with respect to climate change using satellite data”. In: *Applied Water Science* 9.8 (2019). DOI: [10.1007/s13201-019-1067-9](https://doi.org/10.1007/s13201-019-1067-9).
- [38] M. Saunio et al. “The global methane budget 2000-2012”. In: *Earth System Science Data* 8.2 (2016), pp. 697–751. DOI: [10.5194/essd-8-697-2016](https://doi.org/10.5194/essd-8-697-2016).
- [39] M. K. W. Ko et al. *SPARC Report on Lifetimes of Stratospheric Ozone-Depleting Substances, Their Replacements, and Related Species*. Tech. rep. SPARC, 2013, 256 pp. URL: <http://www.sparc-climate.org/publications/sparc-reports/>.
- [40] J. L. Laughner and R. C. Cohen. “Direct observation of changing NO_x lifetime in North American cities”. In: *Science* 366.6466 (2019), pp. 723–727. DOI: [10.1126/science.aax6832](https://doi.org/10.1126/science.aax6832).
- [41] B. Mijling and R. J. van der A. “Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale”. In: *Journal of Geophysical Research: Atmospheres* 117.D17 (2012), n/a–n/a. DOI: [10.1029/2012jd017817](https://doi.org/10.1029/2012jd017817).
- [42] H. S. Kenagy et al. “NO_x Lifetime and NO_y Partitioning During WINTER”. In: *Journal of Geophysical Research: Atmospheres* 123.17 (2018), pp. 9813–9827. DOI: [10.1029/2018jd028736](https://doi.org/10.1029/2018jd028736).
- [43] B. de Foy et al. “Estimates of power plant NO_x emissions and lifetimes from OMI NO₂ satellite retrievals”. In: *Atmospheric Environment* 116 (2015), pp. 1–11. DOI: [10.1016/j.atmosenv.2015.05.056](https://doi.org/10.1016/j.atmosenv.2015.05.056).
- [44] A. Richter. “Nitrogen oxides in the troposphere, What have we learned from satellite measurements?” In: *The European Physical Journal Conferences* 1 (2009), pp. 149–156. DOI: [10.1140/epjconf/e2009-00916-9](https://doi.org/10.1140/epjconf/e2009-00916-9).
- [45] D. Ehhalt et al. “Atmospheric Chemistry and Greenhouse Gases, Chapter 4 of the IPCC Third Assessment Report Climate Change 2001: The Scientific Basis.” In: *IPCC Third Assessment Report Climate Change 2001* (Oct. 2001).
- [46] IPCC. “Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change”. In: *Cambridge University Press*. (2021).
- [47] T. Stavrou et al. “Key chemical NO_x sink uncertainties and how they influence top-down emissions of nitrogen oxides”. In: *Atmospheric Chemistry and Physics* 13.17 (2013), pp. 9057–9082. DOI: [10.5194/acp-13-9057-2013](https://doi.org/10.5194/acp-13-9057-2013).
- [48] T. Holloway, H. Levy, and P. Kasibhatla. “Global distribution of carbon monoxide”. In: *Journal of Geophysical Research: Atmospheres* 105.D10 (2000), pp. 12123–12147. DOI: [10.1029/1999jd901173](https://doi.org/10.1029/1999jd901173).
- [49] M. Saunio et al. “The Global Methane Budget 2000-2017”. In: *Earth System Science Data* 12.3 (2020), pp. 1561–1623. DOI: [10.5194/essd-12-1561-2020](https://doi.org/10.5194/essd-12-1561-2020).

- [50] J. G. M. Barten et al. "Evaluation of nitrogen oxides (NO_x) sources and sinks and ozone production in Colombia and surrounding areas". In: *Atmospheric Chemistry and Physics* 20.15 (2020), pp. 9441–9458. DOI: [10.5194/acp-20-9441-2020](https://doi.org/10.5194/acp-20-9441-2020).
- [51] J. A. Logan. "Nitrogen oxides in the troposphere: Global and regional budgets". In: *Journal of Geophysical Research* 88.C15 (1983), p. 10785. DOI: [10.1029/jc088ic15p10785](https://doi.org/10.1029/jc088ic15p10785).
- [52] K. E. Pickering et al. "Estimates of lightning NO_x production based on OMI NO₂ observations over the Gulf of Mexico". In: *Journal of Geophysical Research: Atmospheres* 121.14 (2016), pp. 8668–8691. DOI: [10.1002/2015jd024179](https://doi.org/10.1002/2015jd024179).
- [53] S. Kirschke et al. "Three decades of global methane sources and sinks". In: *Nature Geoscience* 6.10 (2013), pp. 813–823. DOI: [10.1038/ngeo1955](https://doi.org/10.1038/ngeo1955).
- [54] T. Hanschmann, H. Deneke, and A. Macke. "The effect of sub-pixel clouds on satellite based cloud detection and TOA cloud radiative effect estimations". In: ().
- [55] N. Eguchi and T. Yokota. "Investigation of clear-sky occurrence rate estimated from CALIOP and MODIS observations". In: *Geophysical Research Letters* 35.23 (2008). DOI: [10.1029/2008gl035897](https://doi.org/10.1029/2008gl035897).
- [56] A. Butz et al. "TROPOMI aboard Sentinel-5 Precursor: Prospective performance of CH₄ retrievals for aerosol and cirrus loaded atmospheres". In: *Remote Sensing of Environment* 120 (2012), pp. 267–276. DOI: [10.1016/j.rse.2011.05.030](https://doi.org/10.1016/j.rse.2011.05.030).
- [57] O. Hasekamp et al. *Algorithm Theoretical Baseline Document for Sentinel-5 Precursor Methane Retrieval*. 2019.
- [58] D. J. Varon et al. "Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes". In: *Atmospheric Measurement Techniques* 11.10 (2018), pp. 5673–5686. DOI: [10.5194/amt-11-5673-2018](https://doi.org/10.5194/amt-11-5673-2018).
- [59] J. Olauson. "ERA5: The new champion of wind power modelling?" In: *Renewable Energy* 126 (2018), pp. 322–331. DOI: [10.1016/j.renene.2018.03.056](https://doi.org/10.1016/j.renene.2018.03.056).
- [60] J. Tan et al. "IMERG V06: Changes to the Morphing Algorithm". In: *Journal of Atmospheric and Oceanic Technology* 36.12 (2019), pp. 2471–2482. DOI: [10.1175/jtech-d-19-0114.1](https://doi.org/10.1175/jtech-d-19-0114.1).
- [61] Meteoblue. *ERA5: Technical Documentation*. Last access: 16/05/2022. URL: <https://docs.meteoblue.com/en/meteo/data-sources/era5>.
- [62] A. Molod et al. "Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to Fortuna". In: *Geoscientific Model Development* 8.5 (2012), pp. 1339–1356. DOI: [10.5194/gmd-8-1339-2015](https://doi.org/10.5194/gmd-8-1339-2015).
- [63] M. Crippa et al. "High resolution temporal profiles in the Emissions Database for Global Atmospheric Research". In: *Scientific Data* 7.1 (2020). DOI: [10.1038/s41597-020-0462-2](https://doi.org/10.1038/s41597-020-0462-2).
- [64] J. Kurokawa and T. Ohara. "Long-term historical trends in air pollutant emissions in Asia: Regional Emission inventory in ASia (REAS) version 3". In: *Atmospheric Chemistry and Physics* 20.21 (2020), pp. 12761–12793. DOI: [10.5194/acp-20-12761-2020](https://doi.org/10.5194/acp-20-12761-2020).
- [65] J. J. P. Kuenen et al. "TNO-MACC_II emission inventory; a multi-year (2003-2009) consistent high-resolution European emission inventory for air quality modelling". In: *Atmospheric Chemistry and Physics* 14.20 (2014), pp. 10963–10976. DOI: [10.5194/acp-14-10963-2014](https://doi.org/10.5194/acp-14-10963-2014).
- [66] S. Keita et al. "African Anthropogenic Emissions Inventory for gases and particles from 1990 to 2015". In: *Earth Science Data* (2020). DOI: [10.5194/essd-2020-328](https://doi.org/10.5194/essd-2020-328).
- [67] L. Zhu et al. "Formaldehyde (HCHO) As a Hazardous Air Pollutant: Mapping Surface Air Concentrations from Satellite and Inferring Cancer Risks in the United States". In: *Environmental Science & Technology* 51.10 (2017), pp. 5650–5657. DOI: [10.1021/acs.est.7b01356](https://doi.org/10.1021/acs.est.7b01356).

- [68] H. Chen. “Quantifying the Anthropogenic Methane Emissions from Large Local Sources”. PhD thesis. Aug. 2017.
- [69] P. Cheewaphongphan, S. Chatani, and N. Saigusa. *Exploring Gaps between Bottom-Up and Top-Down Emission Estimates Based on Uncertainties in Multiple Emission Inventories: A Case Study on CH₄ Emissions in China*. 2019. URL: <http://dx.doi.org/10.3390/su11072054>.
- [70] European Environment Agency. *European Industrial Emissions Portal, E-PRTR*. <https://industry.eea.europa.eu/explore/explore-data-map/map>. Last access: 17/12/2021.
- [71] T. Borsdorff et al. “Mapping carbon monoxide pollution from space down to city scales with daily global coverage”. In: *Atmospheric Measurement Techniques* 11.10 (2018), pp. 5507–5518. DOI: [10.5194/amt-11-5507-2018](https://doi.org/10.5194/amt-11-5507-2018).
- [72] L. Wang et al. “Source apportionment of PM 2.5 in top polluted cities in Hebei, China using the CMAQ model”. In: *Atmospheric Environment* 122 (2015), pp. 723–736. DOI: [10.1016/j.atmosenv.2015.10.041](https://doi.org/10.1016/j.atmosenv.2015.10.041).
- [73] Global Energy Monitor. *Global Steel Plant Tracker*. <https://globalenergymonitor.org/projects/global-steel-plant-tracker/>. Last access: 18/12/2021.
- [74] Ministry of Steel Government of India. *Production of Steel*. <https://pib.gov.in/newsite/PrintRelease.aspx?relid=191005>. Last access: 17/12/2021.
- [75] P. Sadavarte and C. Venkataraman. “Trends in multi-pollutant emissions from a technology-linked inventory for India: I. Industry and transport sectors”. In: *Atmospheric Environment* 99 (2014), pp. 353–364. DOI: [10.1016/j.atmosenv.2014.09.081](https://doi.org/10.1016/j.atmosenv.2014.09.081).
- [76] Y. Té et al. “Seasonal variability of surface and column carbon monoxide over the megacity Paris, high-altitude Jungfrauoch and Southern Hemispheric Wollongong stations”. In: *Atmospheric Chemistry and Physics* 16.17 (2016), pp. 10911–10925. DOI: [10.5194/acp-16-10911-2016](https://doi.org/10.5194/acp-16-10911-2016).
- [77] M. Nurunnabi. “The preventive strategies of COVID-19 pandemic in Saudi Arabia”. In: *Journal of Microbiology, Immunology and Infection* 54.1 (2021), pp. 127–128. DOI: [10.1016/j.jmii.2020.07.023](https://doi.org/10.1016/j.jmii.2020.07.023).
- [78] PR Newswire. *CRU: Impact of the Wuhan Coronavirus on the Steel Industry*. <https://www.prnewswire.com/news-releases/cru-impact-of-the-wuhan-coronavirus-on-the-steel-industry-300996817.html>. Last access: 18/12/2021.
- [79] World Steel Association. *Top steel-producing companies*. <https://web.archive.org/web/20160719102615/http://www.worldsteel.org/statistics/top-producers.html>. Last access: 13/05/2021.
- [80] PIME Asia News. *Chinese steel mills start to close following government orders*. <https://www.asianews.it/news-en/Chinese-steel-mills-start-to-close-following-government-orders-38031.html>. Last access: 18/12/2021.
- [81] J. Jiang. *Boasteel to merge with WISCO*. <https://splash247.com/boasteel-merge-wisco/>. Last access: 13/05/2022.
- [82] I. Irakulis-Loitxate et al. “Satellites Detect Abatable Super-Emissions in One of the World’s Largest Methane Hotspot Regions”. In: *Environmental Science & Technology* 56.4 (2022), pp. 2143–2152. DOI: [10.1021/acs.est.1c04873](https://doi.org/10.1021/acs.est.1c04873).
- [83] H. Chen. “Master Thesis: Quantifying the Anthropogenic Methane Emissions from Large Local Sources”. PhD thesis. Aug. 2017.

- [84] D. J. Varon et al. "Satellite Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production". In: *Geophysical Research Letters* 46.22 (2019), pp. 13507–13516. DOI: [10.1029/2019g1083798](https://doi.org/10.1029/2019g1083798).

A

Flowchart for the Divergence Method

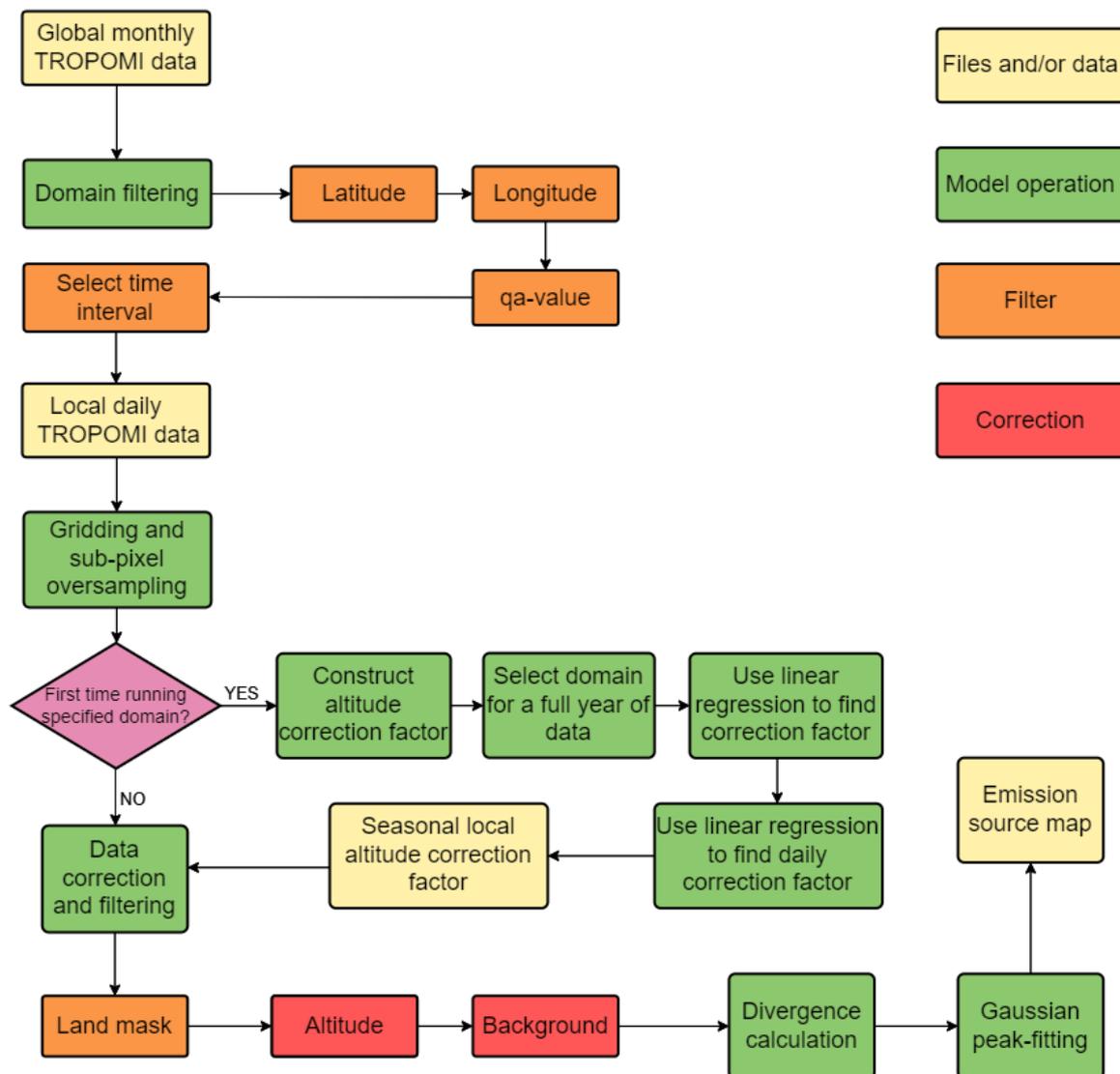


Figure A.1: Flowchart for the divergence method.

B

WRF Pseudo-data Locations

Table B.1: All locations included in the WRF-chem pseudo data. Only non-coastal domains are used for the analysis.

Identifier	Latitude	Longitude	Emissions [Gg/yr]
11	45.437668	27.977934	62.5
13	51.176716	3.8143601	94.2
21	40.516666	17.2	68.5
28	65.56349	22.20601	8.69
30	51.5679	-3.75946	123.0
37	46.943314	18.940529	20.5
42	49.357174	6.754137	101.0
55	52.154766	10.403075	75.0
59	43.45	4.9	74.8
12	52.47656	4.59217	50.8
20	45.61722	13.766944	5.16
27	47.03378	15.067806	71.1
29	53.5812	-0.620021	68.0
36	48.617783	21.198324	114.0
39	43.55611	-5.91113	89.8
45	52.16614	14.617682	38.8
56	53.12494	8.6867285	52.1
61	51.03	2.36	164.0
91	51.50369	6.7359076	117.0
92	51.485516	6.7091227	7.3
93	49.795868	18.305779	50.4
95	50.080334	20.092361	8.66
96	50.34338	19.281864	137.0

C

Case Studies

This section shows the procedure for the case study analysis, the identified emission sources and additional figures that are the result of the case studies.

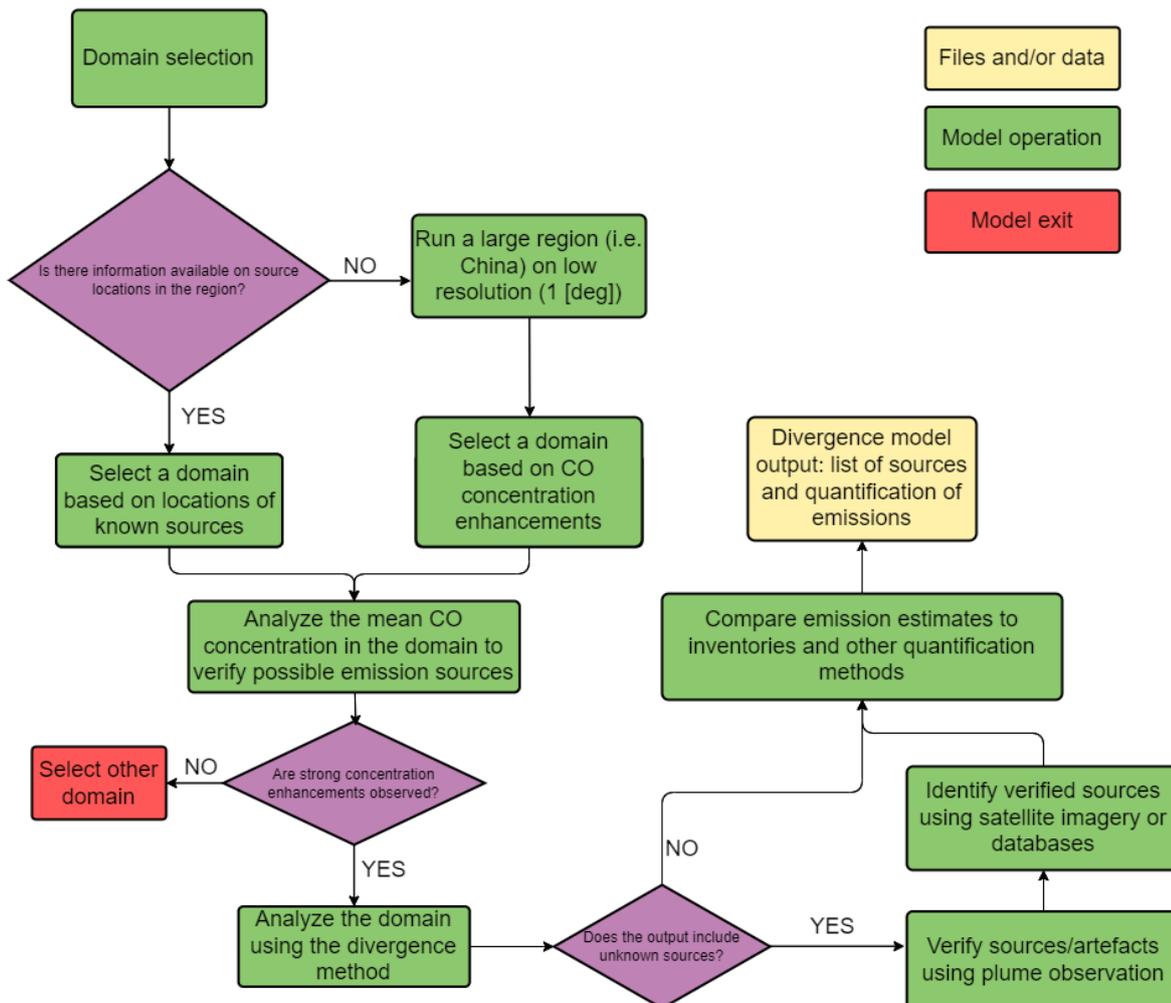


Figure C.1: The case study procedure, showing all necessary steps to study a location.

C.1. Identified sources

Table C.1: Identified emission sources by the divergence method for selected case studies

Case study	Source name	Source type	Lat	Lon	Emission quantification [Gg/a]
Duisburg	Thyssenkrupp N/S	Industrial	51.49	6.73	160 [130-190]
Duisburg	Hittenwerke	Industrial	51.37	6.72	90 [70-110]
Hebei	Qinhuangdao Hongxing	Industrial	39.66	118.89	1,300 [1,000-1,500]
Hebei	Tangshan Donghai	Industrial	39.63	118.42	850 [680-1,000]
Hebei	Qian'an Liang Yanshan	Industrial	39.92	118.68	1,400 [1,100-1,600]
Hebei	Hebei Jinxi	Industrial	40.21	118.22	620 [500-750]
Hebei	Tangshan Ruifeng	Industrial	39.47	118.25	210 [170-250]
Jharkhand	Rourkela	Industrial	22.21	84.87	280 [220 - 330]
Jharkhand	Durgapur	Industrial	23.55	87.25	400 [320-480]
Jharkhand	Bokaro	Industrial	23.68	86.10	390 [310-470]
Jharkhand	IISCO	Industrial	23.67	86.92	260 [210-310]
Jharkhand	TATA	Industrial	22.79	86.20	450 [360-540]
Jharkhand	Metaliks	Industrial	22.38	87.28	120 [100-140]
Jharkhand	Unknown B	Unknown	23.80	86.38	740 [600-900]
Jharkhand	Saluja	Industrial	24.14	86.34	190 [150-230]
Cairo	Cairo	Urban	30.04	31.24	520 [420-620]
Riyadh	Riyadh	Urban	24.61	46.72	250 [200-300]
Wuhan	Qingshan	Industrial	30.64	114.46	540 [430-650] (2019) 450 [360-540] (2020)
Wuhan	Echeng	Industrial	30.40	114.86	270 [220-320] (2019) 290 [230-350] (2020)
Wuhan	Huangshi	Industrial	30.07	114.92	130 [100-160] (2019) 100 [80-120] (2020)

C.2. Duisburg, Germany

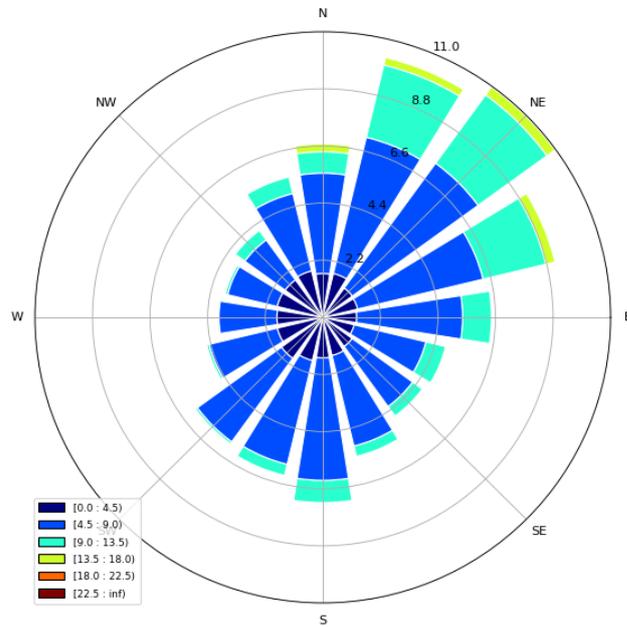


Figure C.2: Wind rose for the Duisburg area, showing wind direction and speed.

Absolute wind speed for Duisburg on 20180701 - 20210630
Grid resolution: 0.03

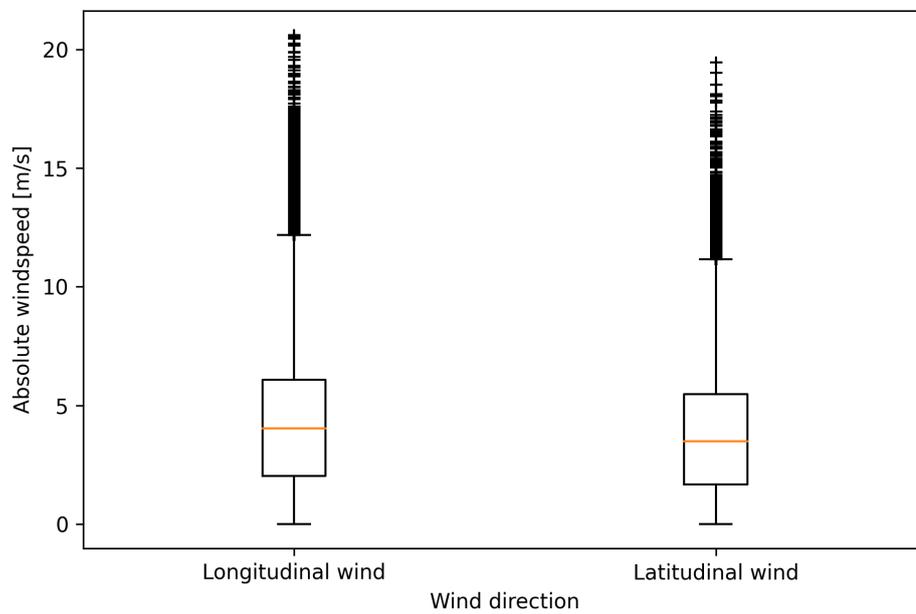


Figure C.3: The different latitudinal and longitudinal wind speeds for the Duisburg area. The '+' markers indicate outliers.

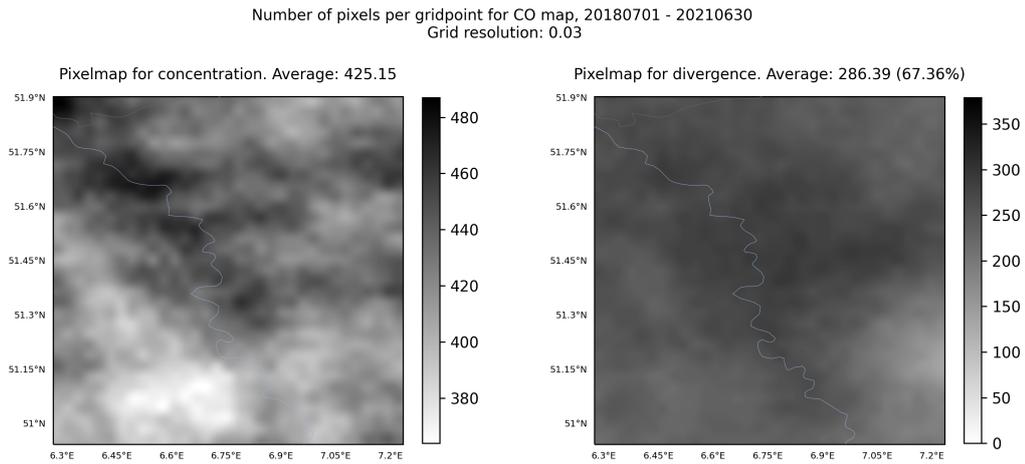


Figure C.4: Average grid coverage for the Duisburg area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.

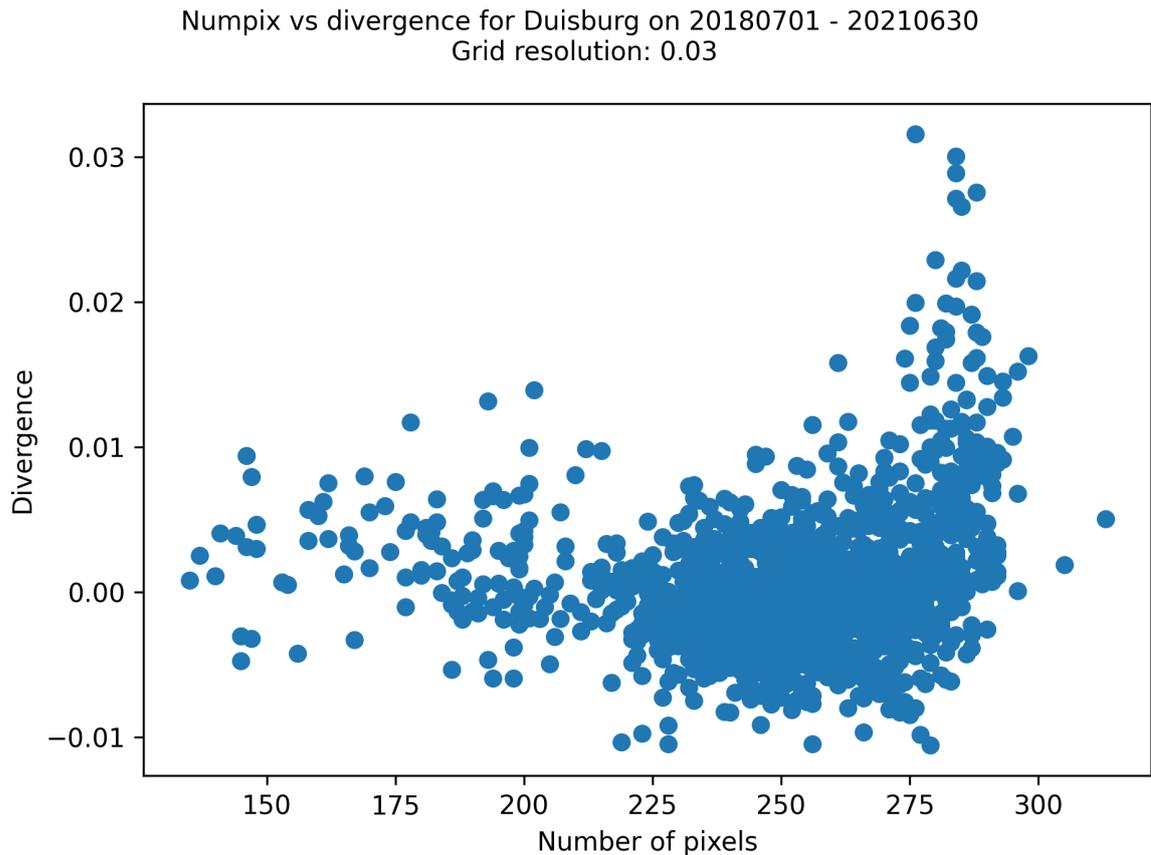


Figure C.5: Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact. No artefacts are observed.

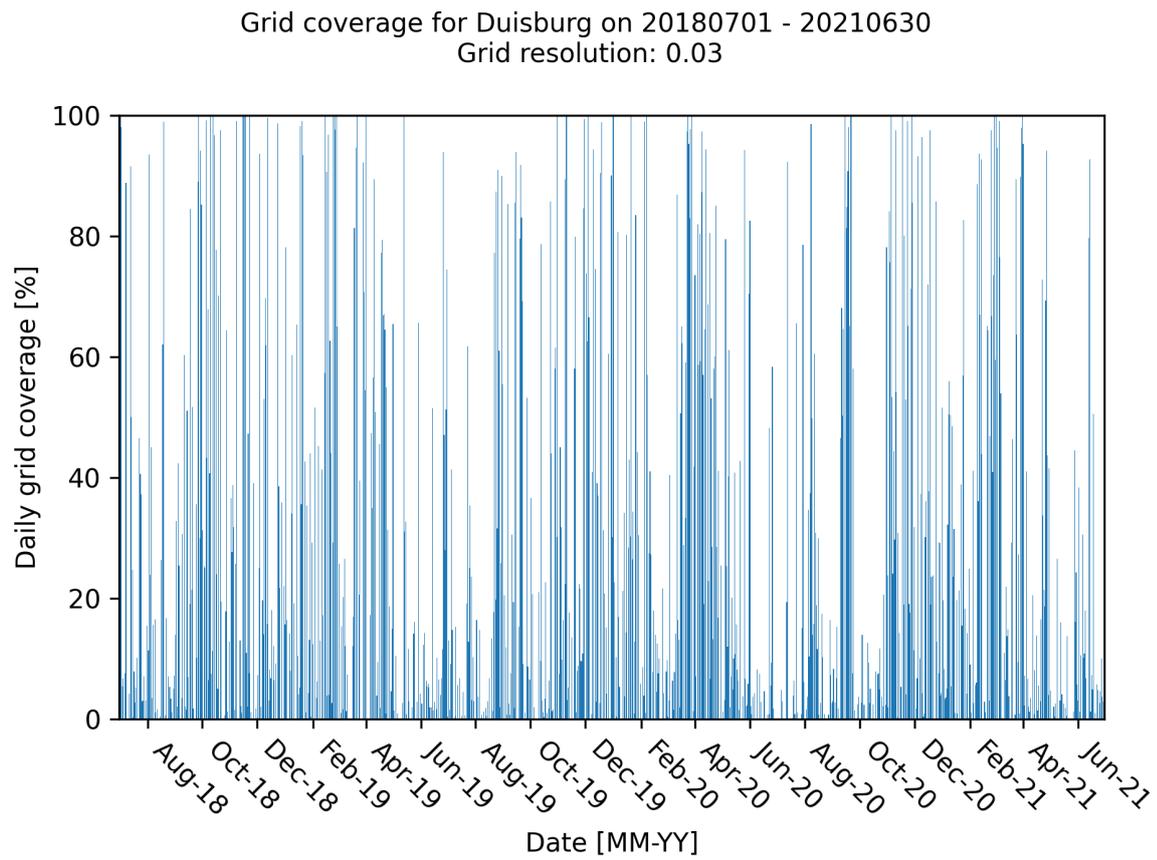


Figure C.6: Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The domain has reasonable coverage year-round.

C.3. Hebei, China

CO concentration map, 20180701 - 20210630

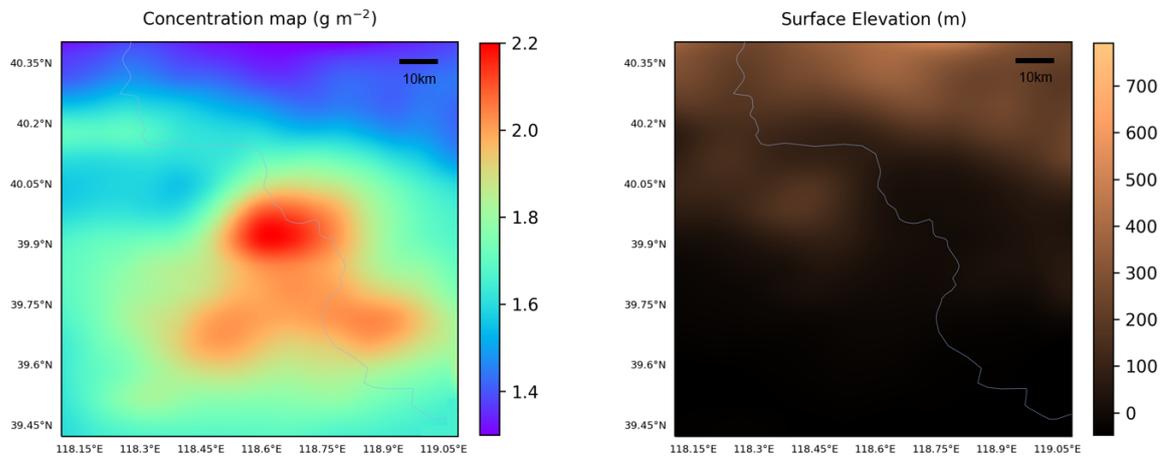


Figure C.7: The Hebei area, showing the TROPOMI CO concentration (left) and the surface elevation (right) at low resolution (0.07°). Multiple concentration enhancements can be spotted.

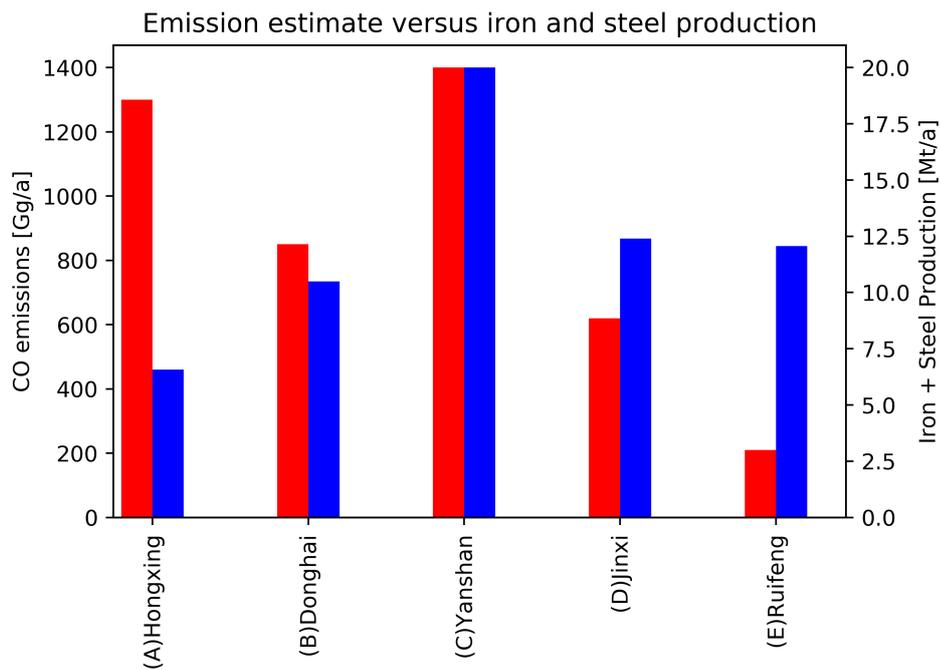


Figure C.8: Emission estimates (red) from the divergence method for the five plants in Hebei versus the crude iron and steel production for 2020 (blue) [73].

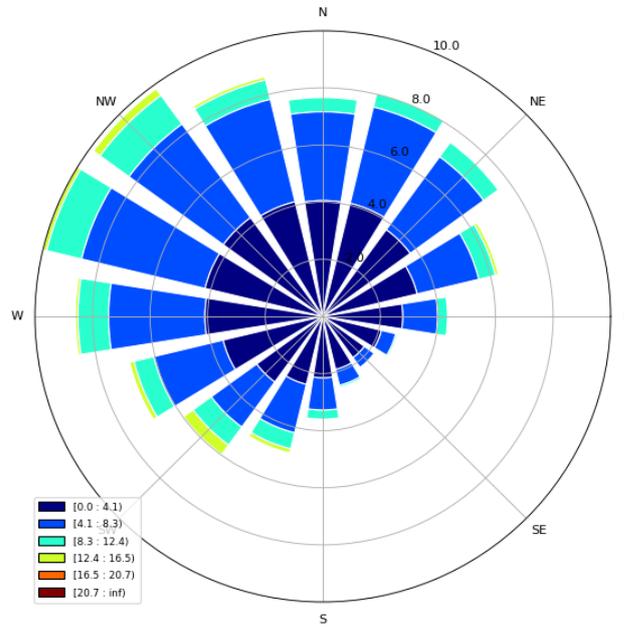


Figure C.9: Wind rose for the Hebei area, showing wind direction and speed.

Absolute wind speed for Nanpi on 20180701 - 20210630
Grid resolution: 0.03

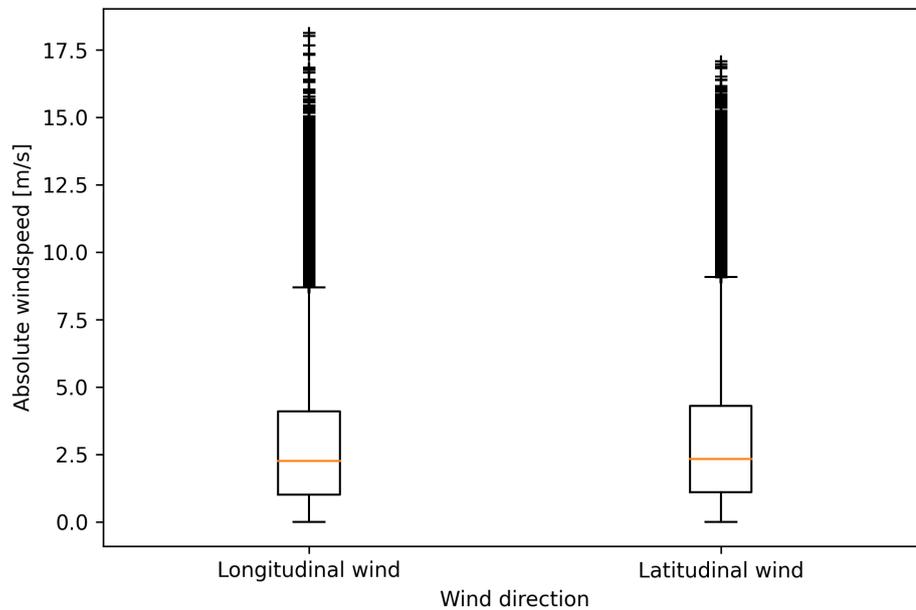


Figure C.10: The different latitudinal and longitudinal wind speeds for the Hebei area. The '+' markers indicate outliers.

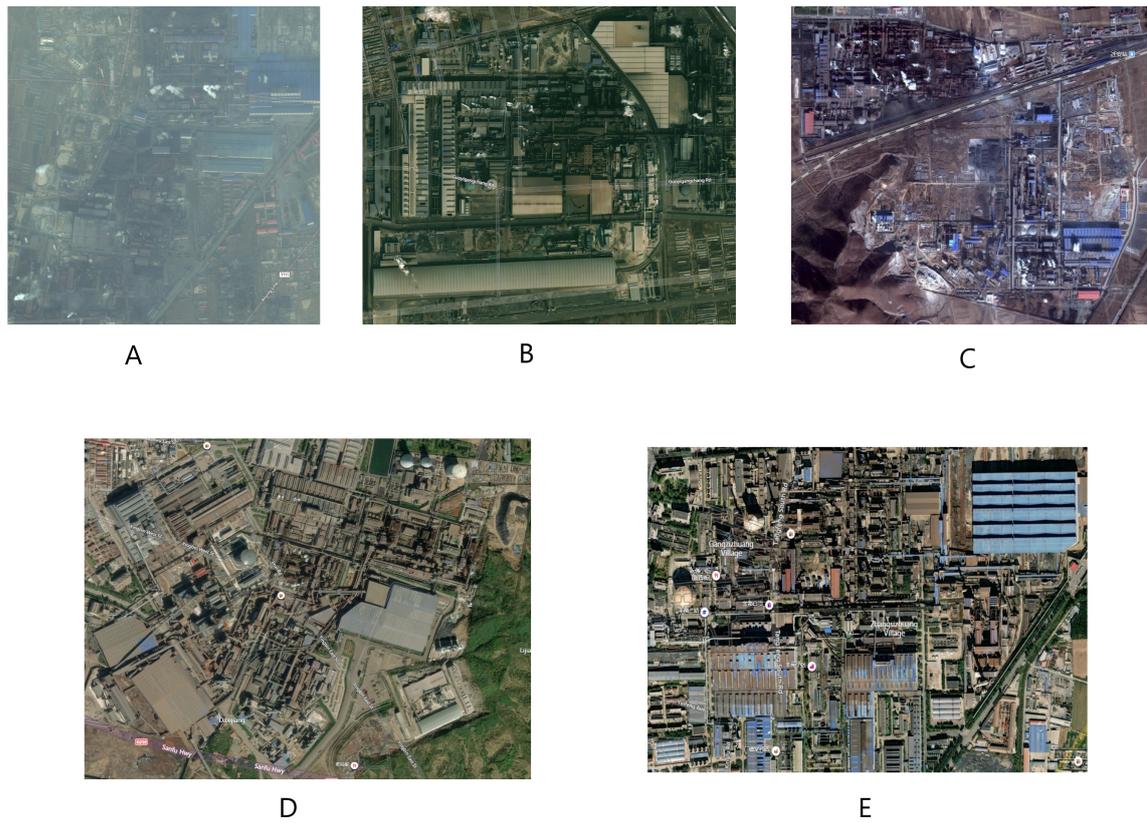


Figure C.11: The five resulting sources from the Hebei model run. All location match a steel/power plant in the Hebei area. Images are taken from Bing maps.

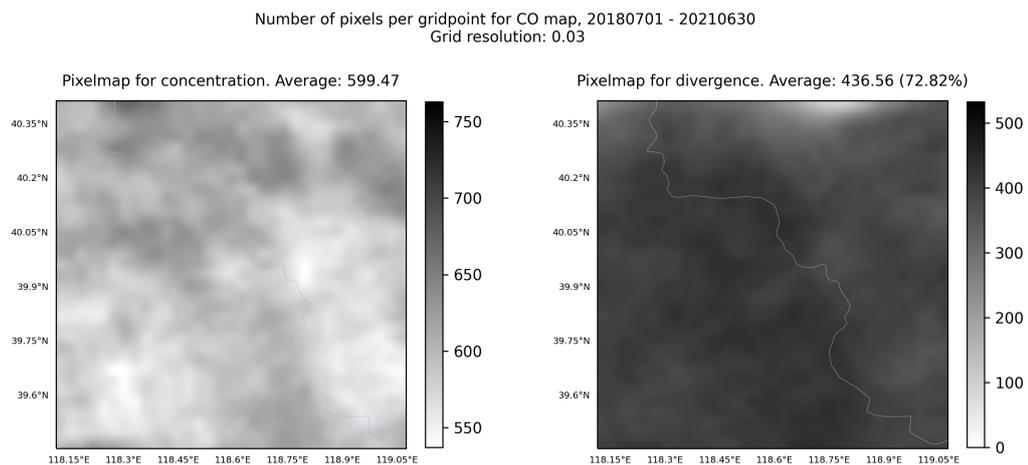


Figure C.12: Average grid coverage for the Hebei area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.

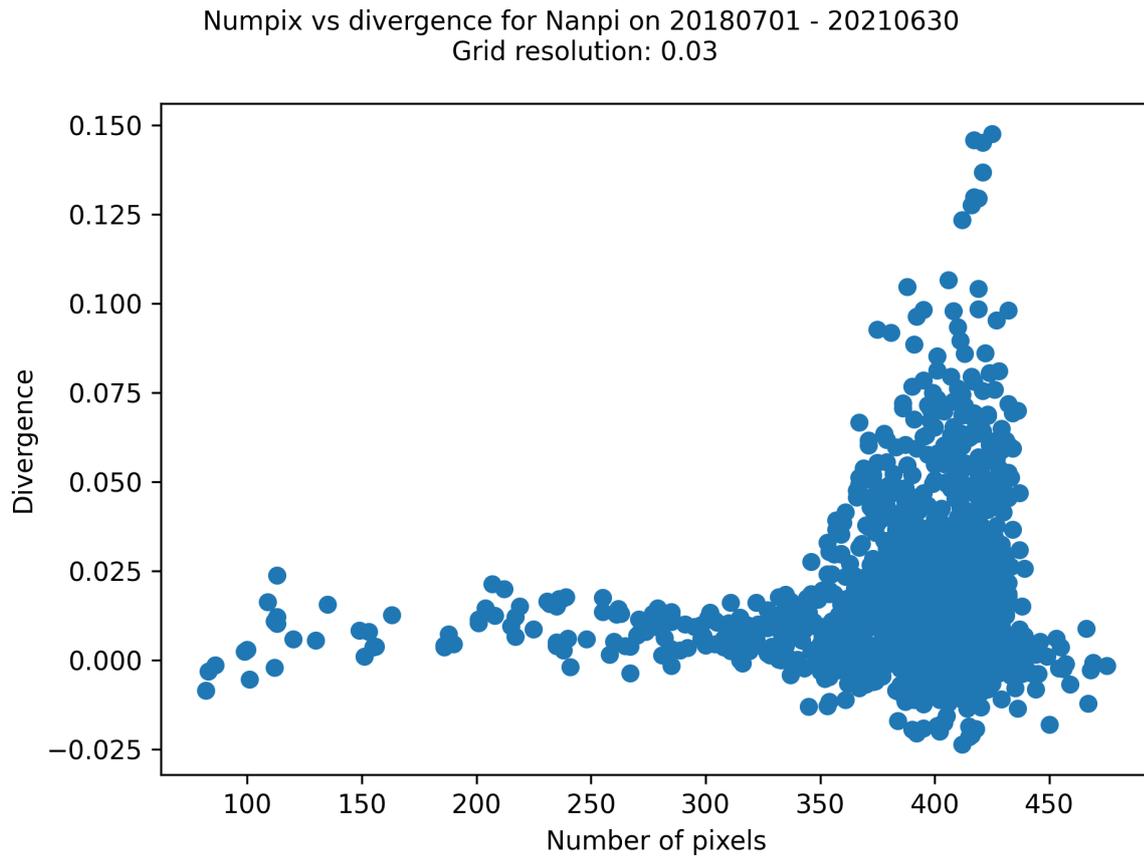


Figure C.13: Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact. No artefacts are identified.

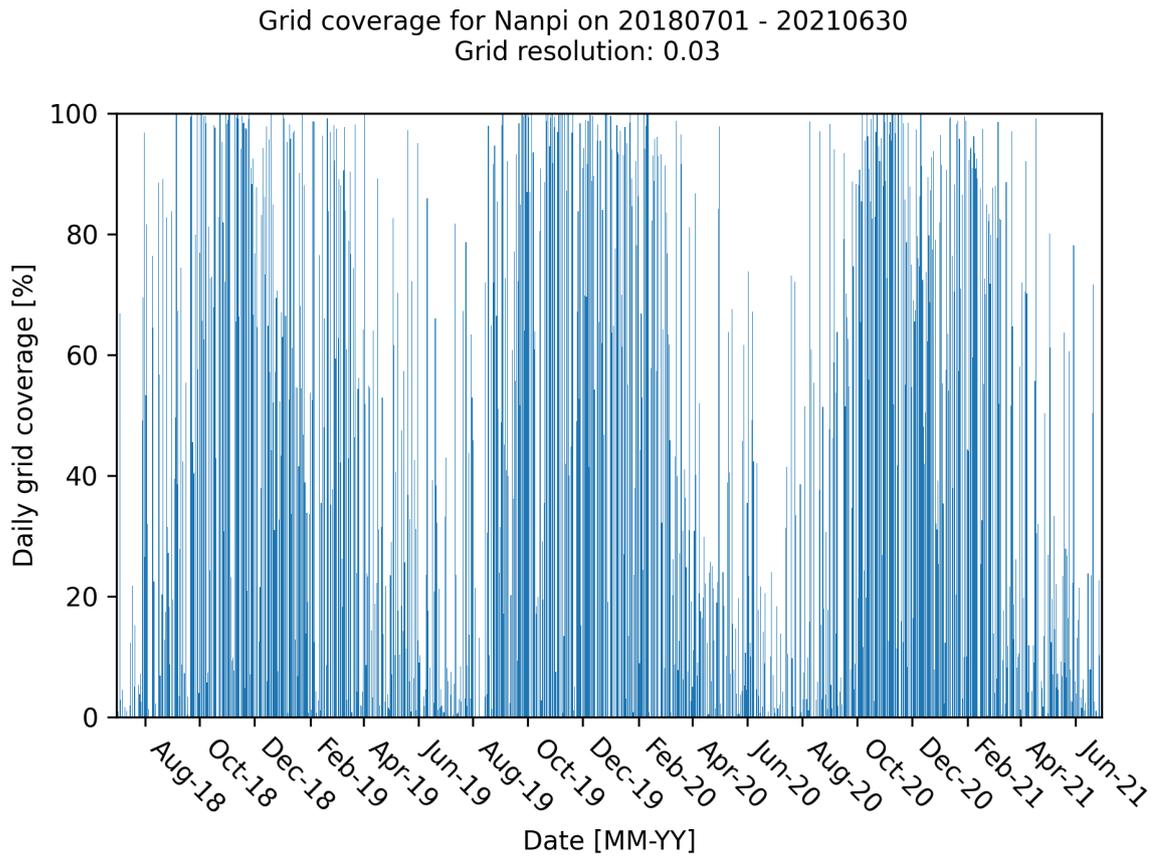


Figure C.14: Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The monsoon season with high cloud coverage can be clearly seen during the summer months.

C.4. Jharkhand, India

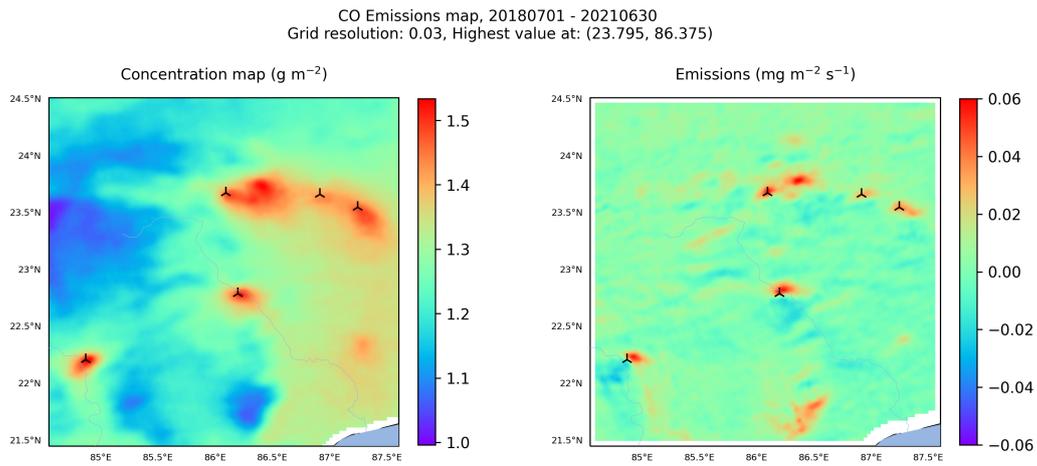


Figure C.15: The Jharkhand area, showing the CO concentration (left) and the divergence (right) at high resolution (0.03°) without altitude correction. The markers indicate the locations of known emission sources. Multiple altitude-induced artefacts can be spotted.

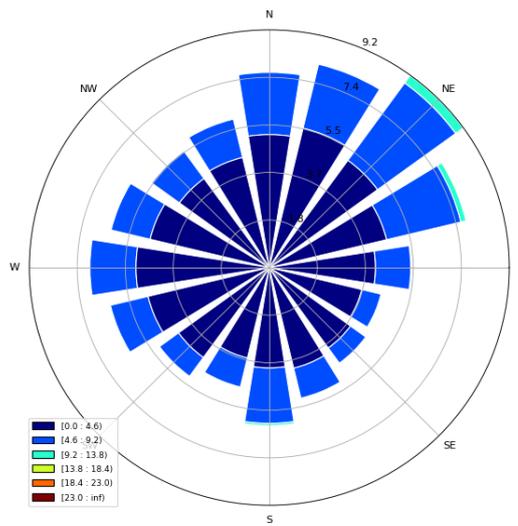


Figure C.16: Wind rose for the Jharkhand area, showing wind direction and speed.

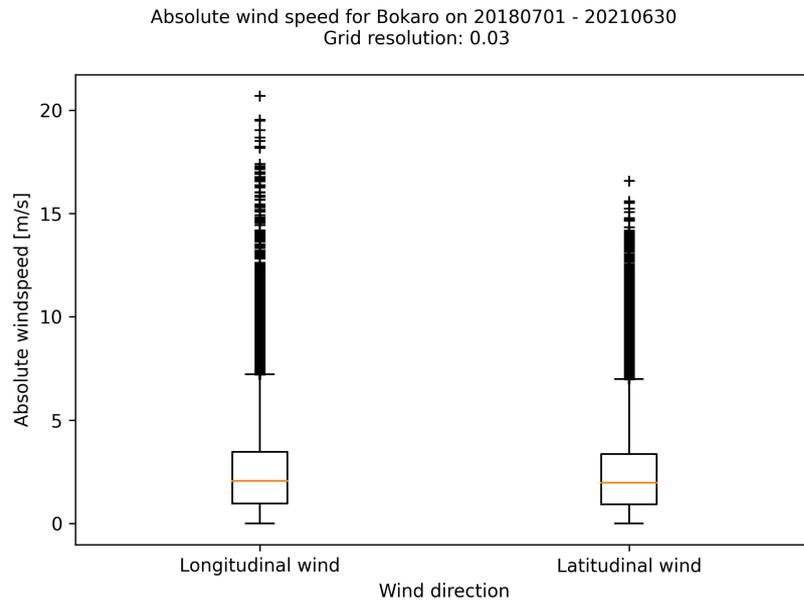


Figure C.17: The different latitudinal and longitudinal wind speeds for the Jharkhand area. The '+' markers indicate outliers.

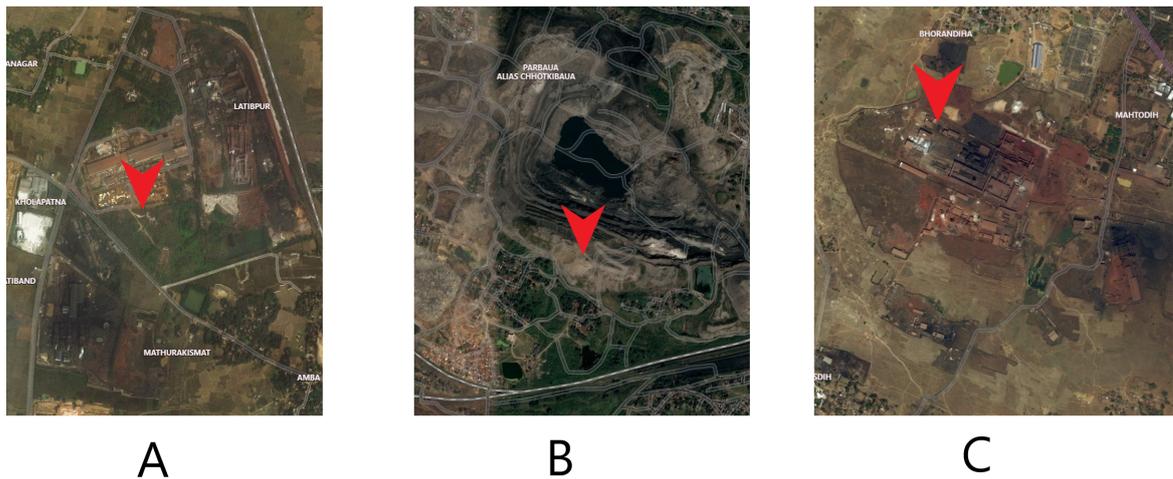


Figure C.18: The two unknown locations from the Jharkhand model run. (A) is identified as the Metaliks cement and power plant. (B) is still unidentified. (C) is identified as the Saluja steel plant. Images are taken from Bing maps.

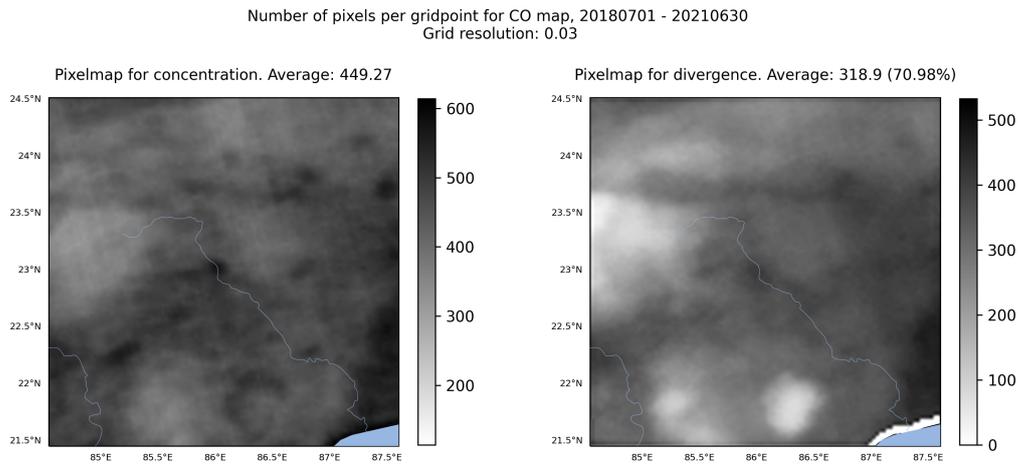


Figure C.19: Average grid coverage for the Jharkhand area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.

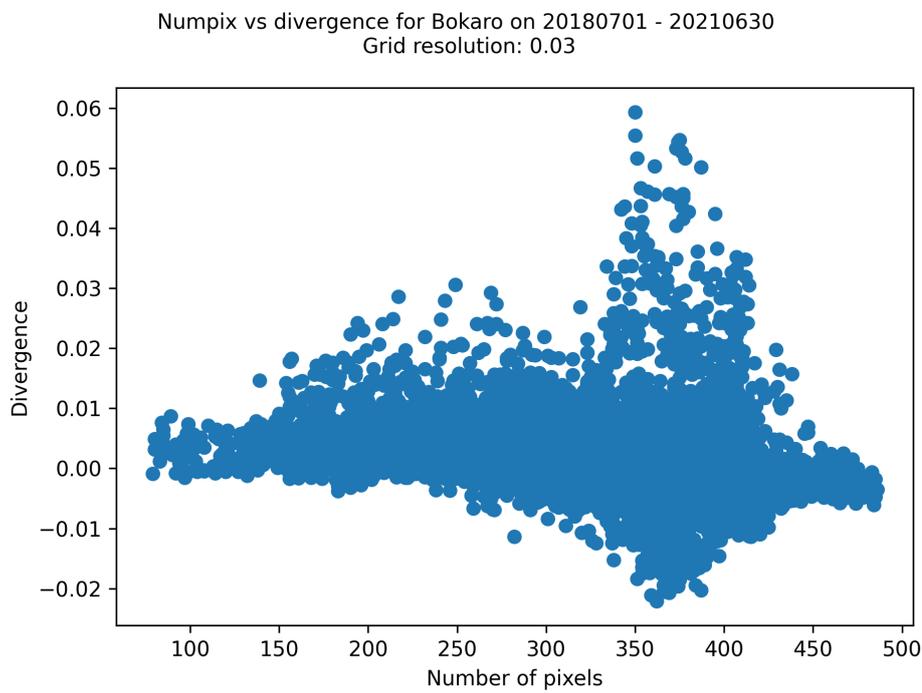


Figure C.20: Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact. No artefact indicates are observed.

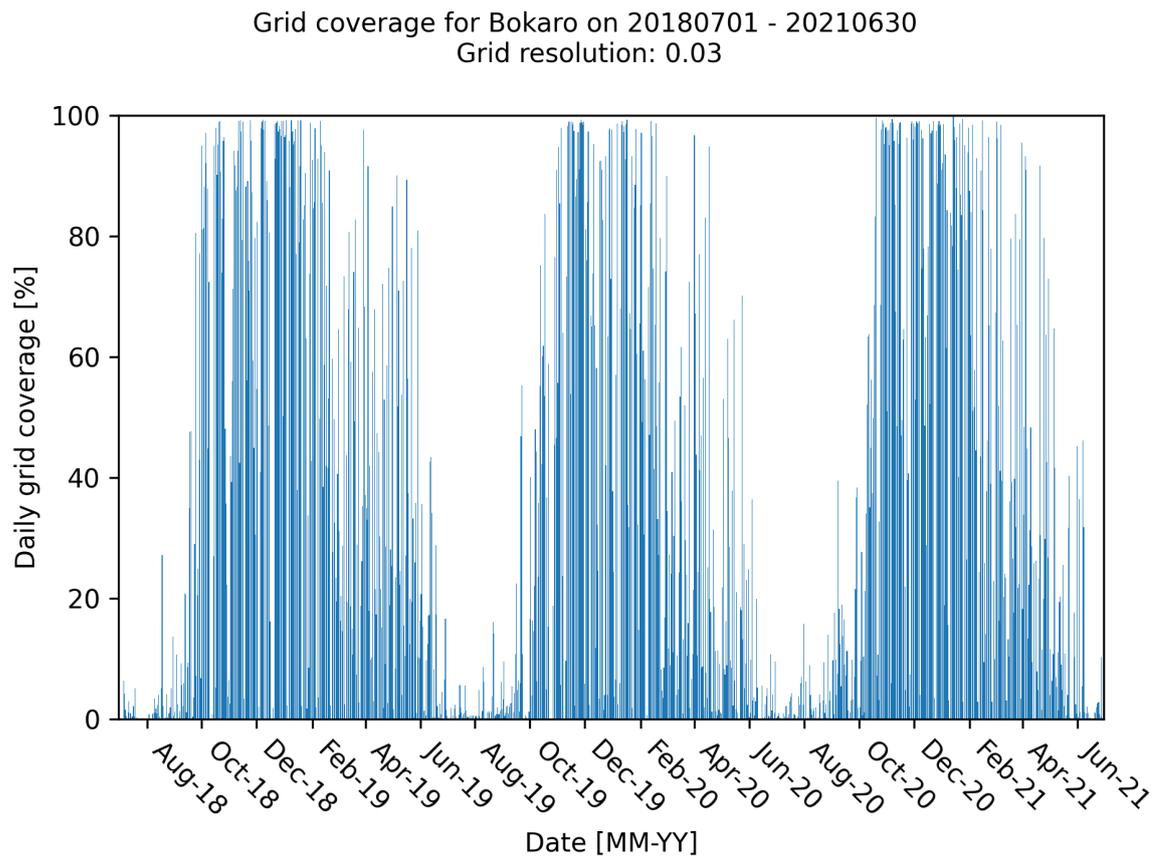


Figure C.21: Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The monsoon season with high cloud coverage can be clearly seen during the summer months.

C.5. Cairo, Egypt

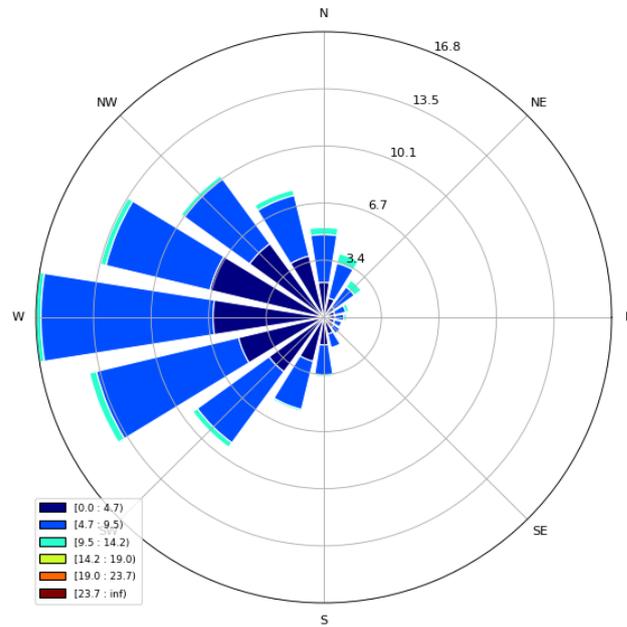


Figure C.22: Wind rose for the Cairo area, showing wind direction and speed.

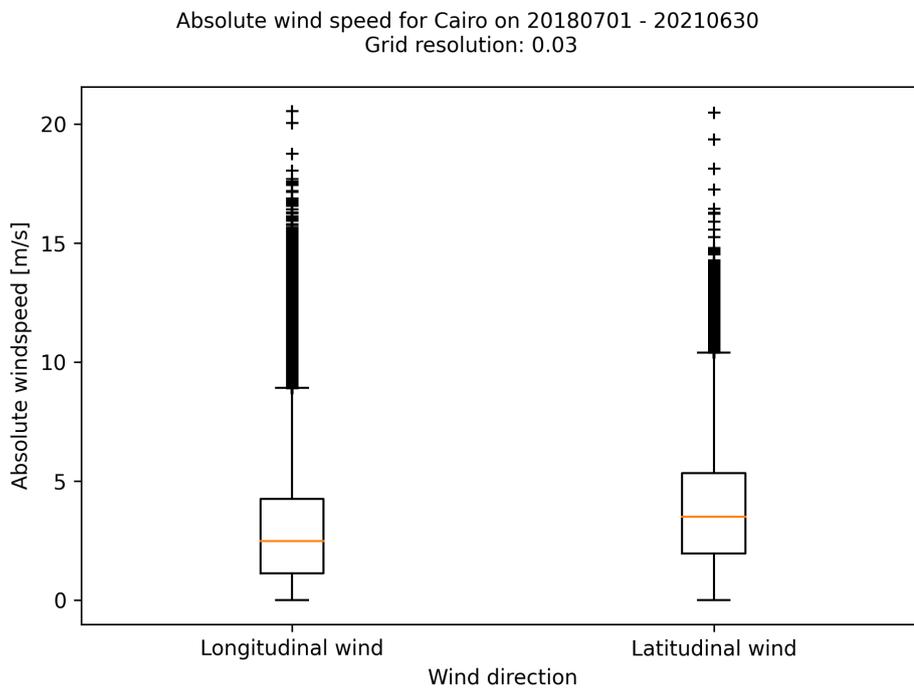


Figure C.23: The different latitudinal and longitudinal wind speeds for the Cairo area. The '+' markers indicate outliers.

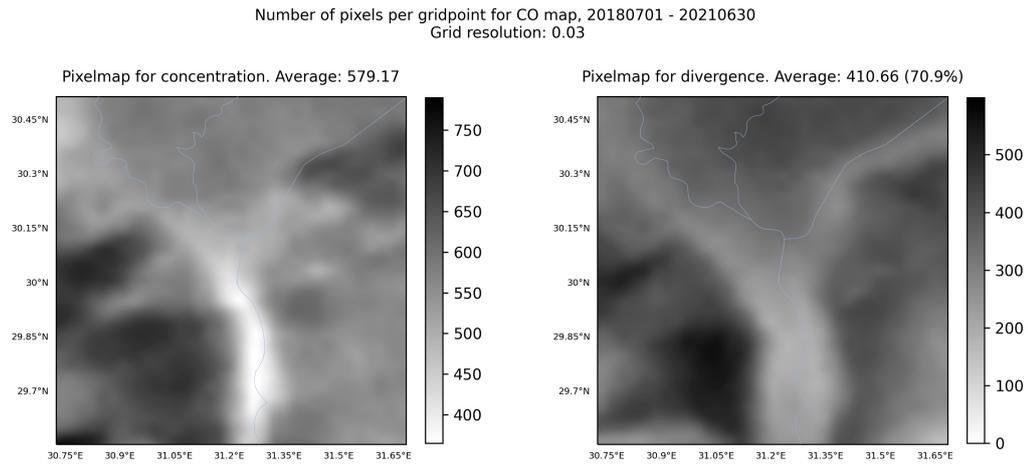


Figure C.24: Average grid coverage for the Cairo area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.

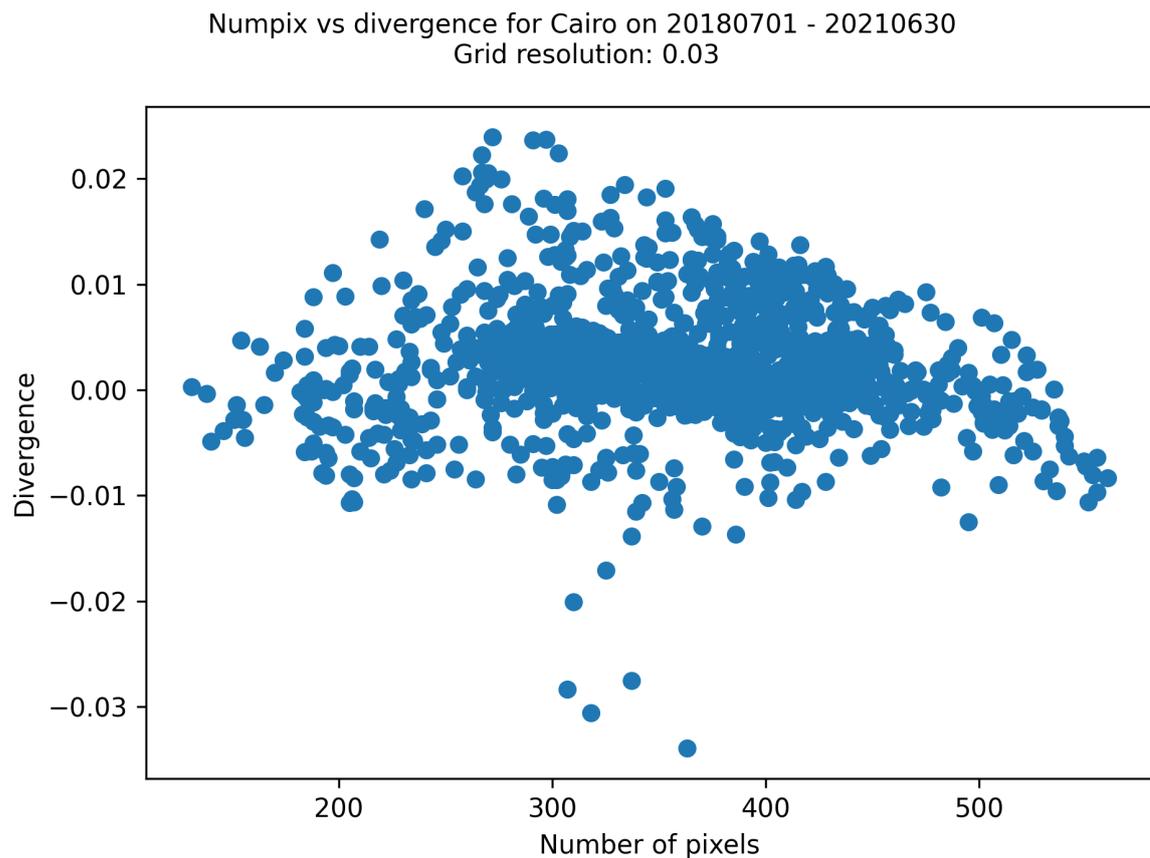


Figure C.25: Divergence and numpix values for every gridpoint. The isolated large negative divergence values around 300-350 might indicate an artefact.



Figure C.26: The location of the artefact, as identified by the divergence method.

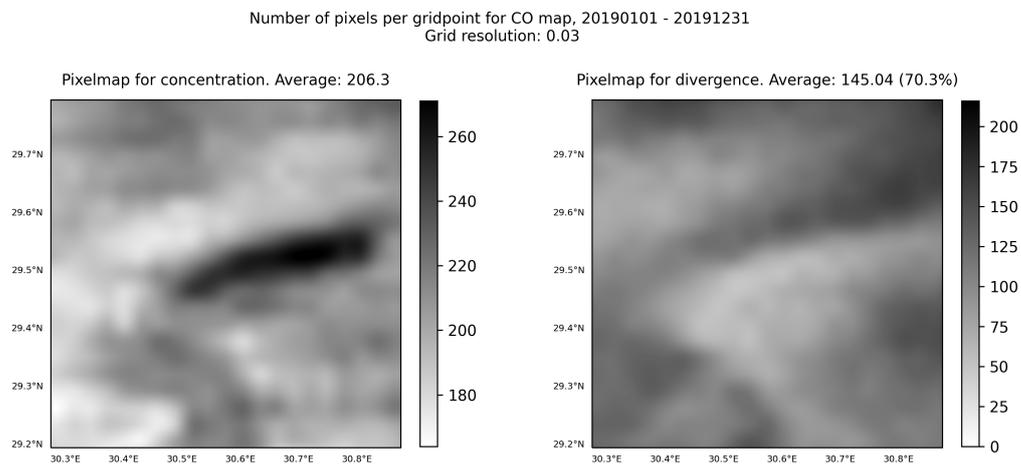


Figure C.27: The numpix map for the artefact location. A clear numpix enhancement can be seen in the center of the grid.

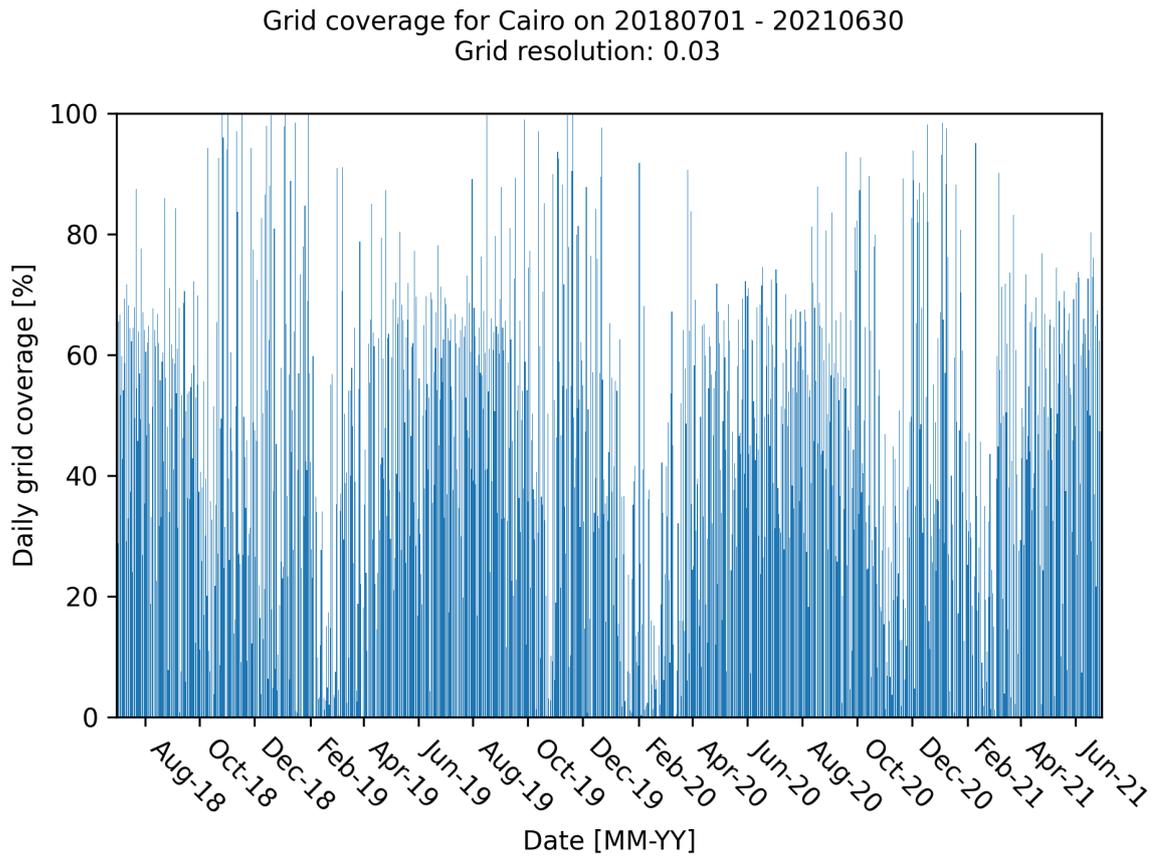


Figure C.28: Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The grid has low coverage during the winter months, with high coverage in the summer.

C.6. Riyadh, Saudi-Arabia

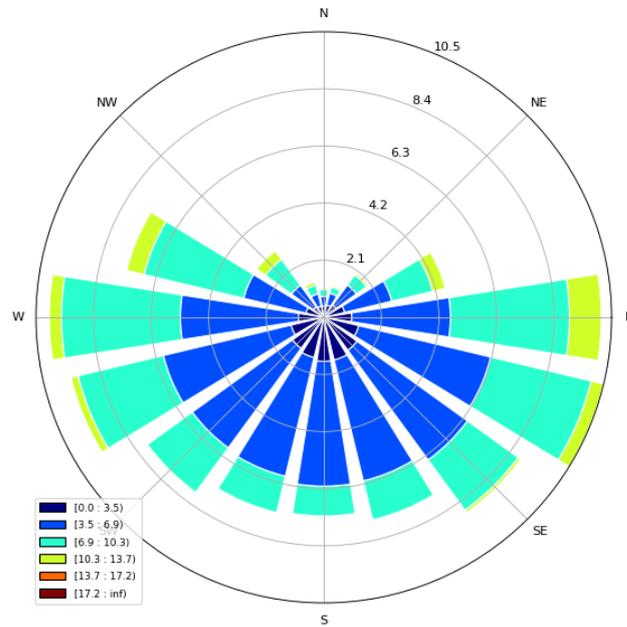


Figure C.29: Wind rose for the Riyadh area, showing wind direction and speed.

Absolute wind speed for Riyadh on 20180701 - 20210630
Grid resolution: 0.03

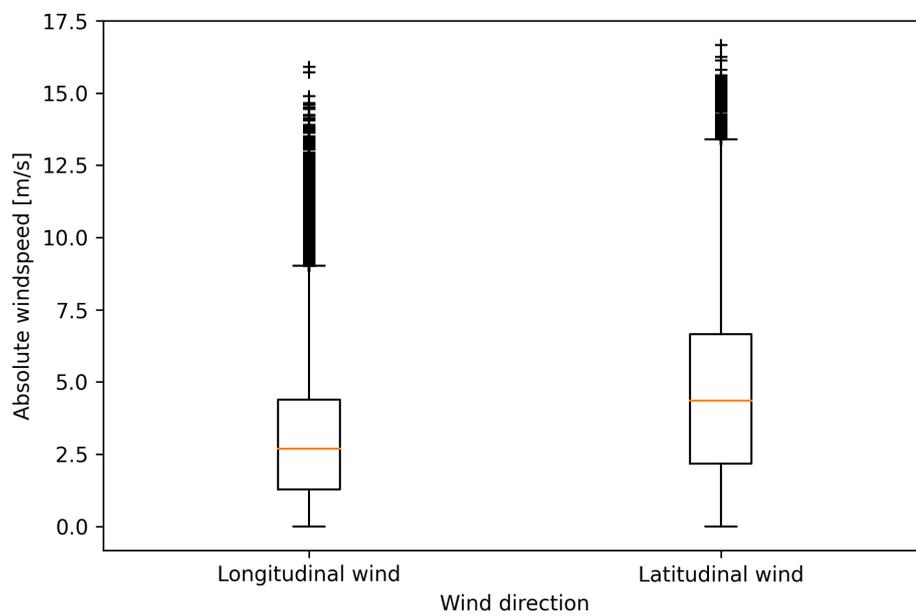


Figure C.30: The different latitudinal and longitudinal wind speeds for the Riyadh area. The '+' markers indicate outliers.

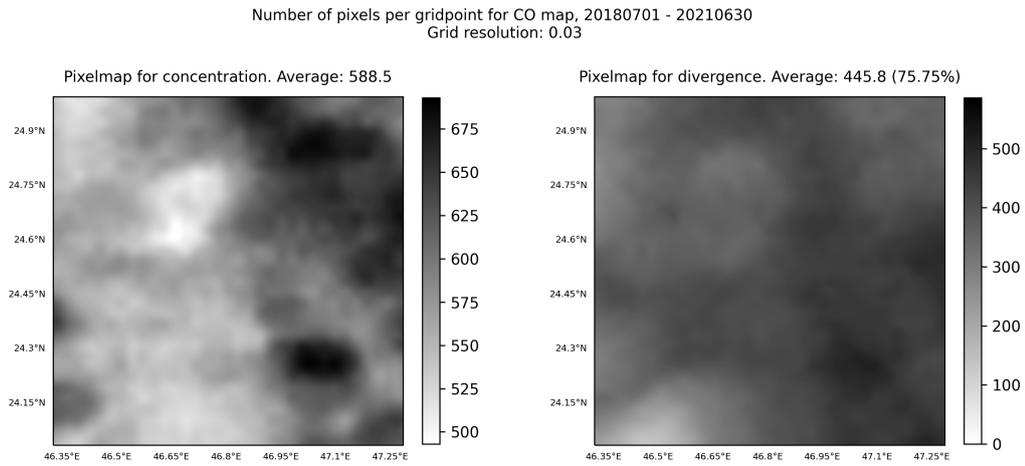


Figure C.31: Average grid coverage for the Riyadh area. The coverage is shown as the average TROPOMI pixels used for every gridcell for both the concentration and the divergence. The percentage above the divergence pixelmap indicates the data loss due to the numerical derivative.

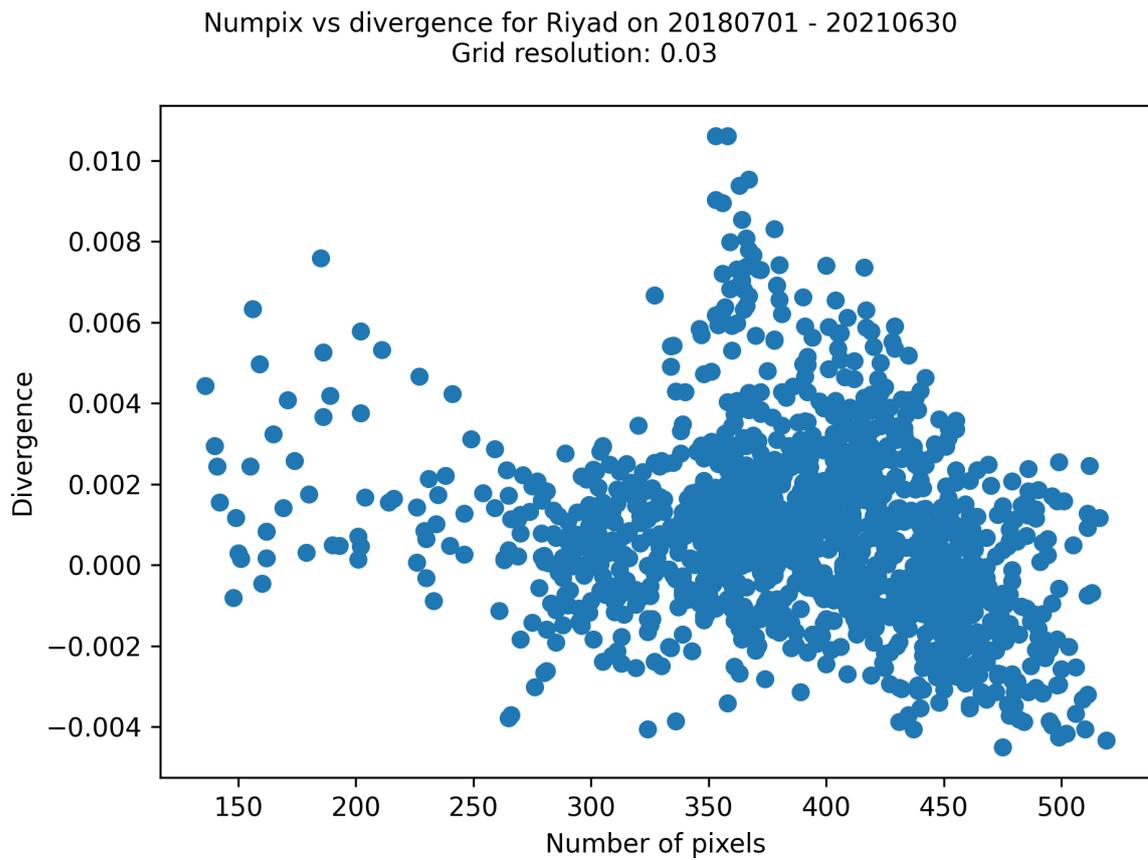


Figure C.32: Divergence and numpix values for every gridpoint. High divergence at low numpix indicates an artefact.

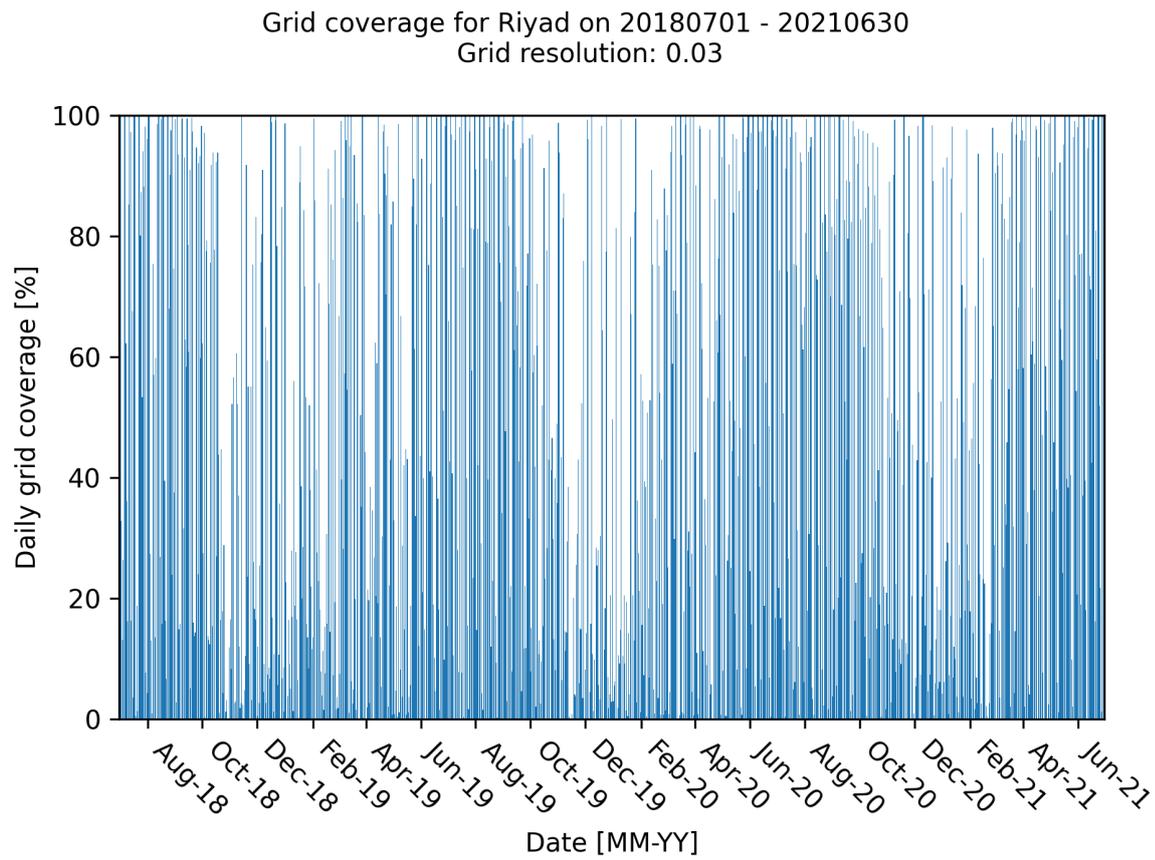


Figure C.33: Average grid coverage for the full time domain. The coverage is a percentage of the gridpoints with a value for the concentration. The domain has good coverage year-round.

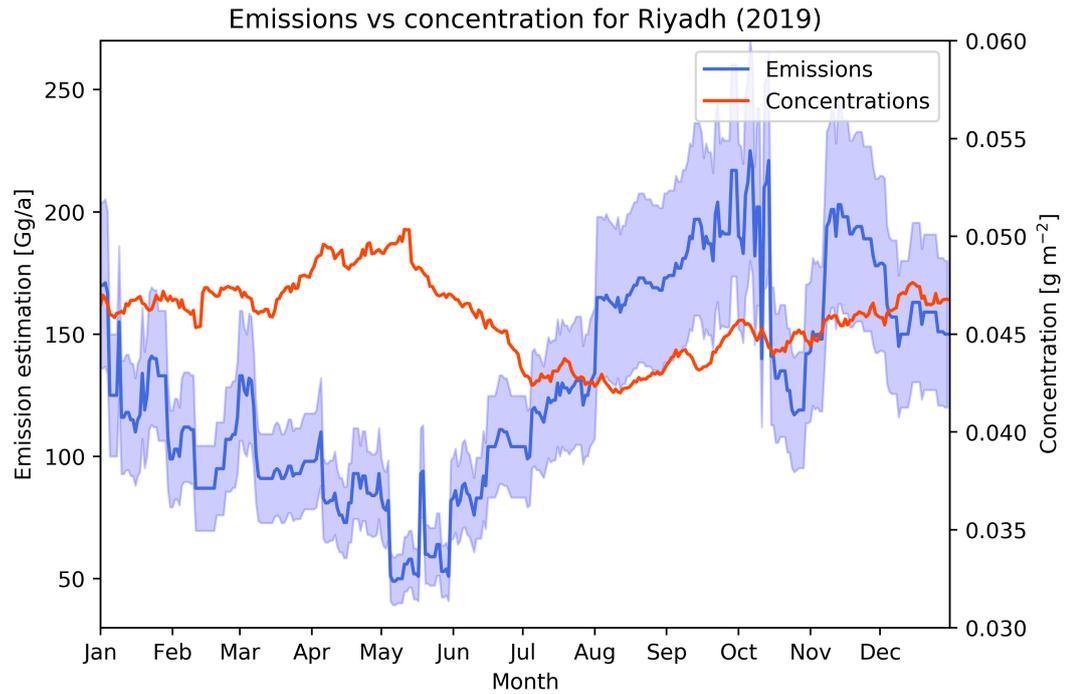


Figure C.34: Mean background-corrected CO concentration (red) and estimated CO emissions (blue) for Riyadh in 2019 using a 90-day running mean. The blue band shows the uncertainty for the emission estimation.

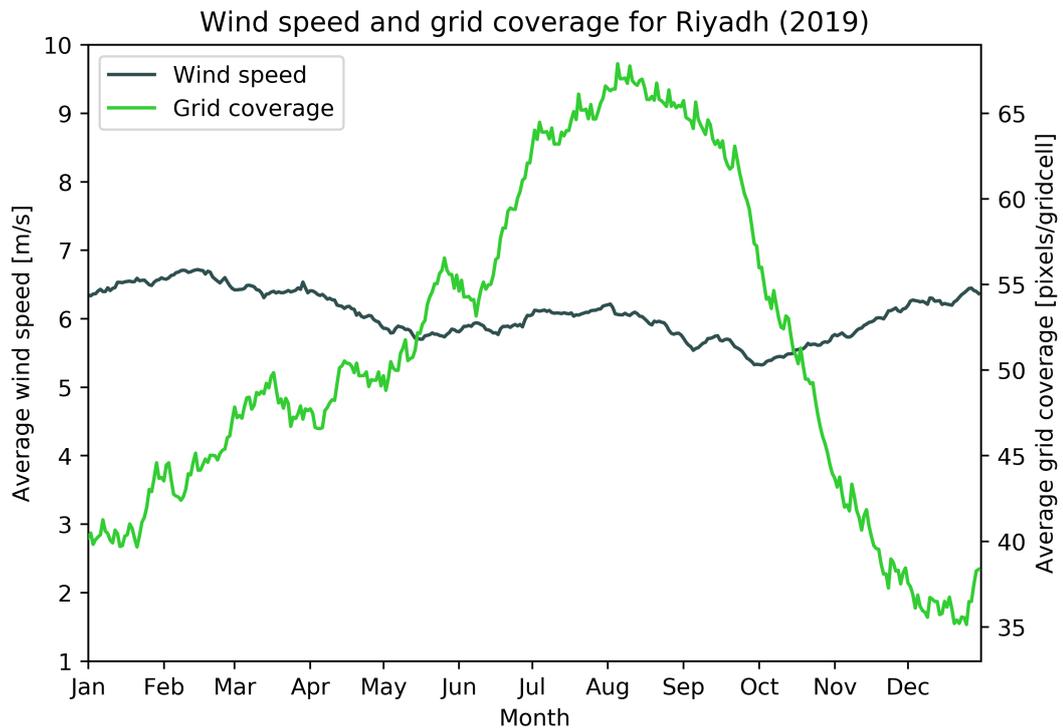


Figure C.35: The average wind speeds and grid coverage for the Riyadh 2019 analysis. A strong seasonality for grid coverage can be seen. The average wind speed only slightly differs.