We would like to thank reviewer #1 for the helpful review and revised the manuscript based on their comments. Please find our detailed response and the revised manuscript below. Please also note the response to reviewer #2.

**General comments**

**1. Reviewer #1:** The proposed method cannot be applied to real satellite data as it is now, because a key problem is not addressed, namely how to derive the background and its uncertainty. This should be discussed more prominently also in the conclusions and already be mentioned in the abstract. To me, many points in the description of the algorithm are unclear (see my specific comments).

**Response:** We revised and restructured the description of our algorithm (Section 2.2.1) making everything hopefully more clear.

In our study, we estimate background and its variance from the corresponding component in the COSMO-GHG simulations. We think that this approach is sufficiently accurate for comparing the detectability of $CO_2$ plumes with $CO_2$, $NO_2$ and CO observations, which was the main objective of this paper. We already discuss the effect of "this optimistic assumption" on the number of detectable plumes in Section 4.3 (Limitations of current plume detection algorithm). We further modified Section 4.3 to clarify our main objective:

>>> *The main objective of this study was to compare the basic detectability of $CO_2$ plumes with $CO_2$, $NO_2$ and CO observations. Our plume detection algorithm was able to detect* weak signals ... <<<

We also replaced "new algorithm" and "novel algorithm" with "simple algorithm" in the manuscript for the same reason. We agree that further development of the plume detection algorithm would be required for its application to real observations. In particular, the background would need to be derived differently. For example, it could be taken from a model or estimated directly from the satellite observations (see Section 2.2.1). We already discuss these options briefly in Section 4.3 and extended the section to provide more details:

>>> *The algorithm assumed accurate knowledge of the mean and variance of the background, which were estimated directly from a simulated background tracer. [...]*

*When applied to real satellite observations the background and its variance could either be taken from a model or estimated directly from the satellite observations. Although atmospheric transport models have large uncertainties at the level of individual plumes, they could provide reasonable estimates of the $CO_2$ and $NO_2$ background. Likely the best option is to derive the background directly from the satellite data, which requires further development of the plume detection algorithm presented here. An improved algorithm could start with an a priori estimate of plume location and background and would then be updated iteratively to improve both plume location and background. A model could be used here to determine a suitable a priori plume location and background. An improved version of the algorithm presented in this paper has the potential for increasing the number of detectable plumes per satellite as well as the number of $CO_2$ pixels per plume.* <<<

**2. Reviewer #1**: NO2 can be used in more cloud contaminated scenes compared to CO2. This is one reason why you find more plumes in NO2. When computing the number of successfully found inner-plume soundings with NO2, do you apply the stricter CO2 cloud filter in the last step? This would be very important, because only these soundings could be used for the flux estimation. Please clarify this in the manuscript and if this last filter is not applied, discuss why you think that your results are still meaningful for CO2M.

**Response**: Yes, we do apply the stricter CO2 cloud filter in the last step. The TP and FP values are thus only computed for cloud-free CO2 pixels. We added the following sentence to Section 2.2.2 (Performance evaluation):

>>>*To remove the impact of different cloud thresholds, TP, FP and PPV were only computed for cloud-free pixels using a cloud threshold of 1%, because we are primarily interested in valid CO2 observations that can be used for estimating CO2 emissions."* <<<


**3. Reviewer #1**: Which effect is dominating the advantages of NO2, less strict cloud filtering or better SNR? How would both compare if CO2 cloud filtering would have been also applied to NO2?

**Response**: Since we did not apply the plume detection to NO2 observations with a 1% cloud threshold, we cannot answer the question quantitatively. We expect that even with a 1% cloud threshold the NO2-based plume detection would still perform significantly better because of the higher signal-to-noise ratio (see Table 6).


**4. Reviewer #1**: In respect to a potential future application to real data, the pros and cons of plume detection based flux retrieval methods should be compared to inverse modeling approaches in more depth. Some examples: Inverse modeling can be applied also to plumes below the detection limit. Inverse modelling may place the plumes wrongly so that modelled plumes become less correlated with measured plumes resulting in underestimated fluxes. Plume detection depends on plume strength, so that this may introduce an observational bias towards large fluxes, i.e., annual average fluxes will likely be biased high, so that comparisons of reported and measured fluxes are more meaningful on a per-overflight basis. Plume detection methods can be applied to unknown source positions. Etc.

**Response**: We would like to thank the reviewer for these insightful points. We included some of the more general points in the introduction and also updates Section 4.3 adding some more specific points:

>>> *The system would allow for observing CO2 plumes of individual point sources such as large cities and power plants and for quantifying the respective emissions during single satellite overpasses (Bovensmann et al. 2010; Pillai et al. 2016, Velazco et al. 2011). A CO2 plume is defined here as an enhancement of CO2 concentrations above the background in the satellite image caused by the emissions of a given source. The emissions of the source can be estimated from the CO2 enhancement inside the plume, which requires that the plume location is identified in the satellite observations. An atmospheric transport model may be used for simulating the plume location and for estimating the emissions with an inversion framework (e.g. Pillai et al. 2016; Broquet et al. 2018). However, the simulated plume might be*

*significantly displaced due to uncertainties in wind fields and emission heights, which would result in systematic errors in the estimated emissions (Broquet et al. 2018, Brunner et al. 2019). It is therefore desirable to detect the plume directly in the satellite observations, which would make it possible to correct transport-related errors in the simulations but also to estimate the emissions directly from the $CO_2$ enhancements in the plume using plume fitting or mass balance approaches, which only require an estimate of the mean wind speed within the plume {Fioletov et al. 2015; Krings et al. 2013; Varon et al 2018}. While the potential for detecting and estimating emissions from $CO_2$ fluxes has been demonstrated for strong $CO_2$ plumes of megacities and large point sources using* the Orbiting Carbon Observatory 2 (OCO-2, Crisp et al. 2017) (Nassar et al 2017, Reuter et al 2019), a major challenge is... <<<

**5. Reviewer #1**: Why do you consider PPV a good measure for performance? As an example: consider a very conservative plume detection algorithm, detecting only soundings which are 20ppm larger than the BG. In this case, TP would admittedly become small, but FP would become zero, so that PPV would become 1. Could the Hanssen-Kuiper Skill Score or True Skill Statistic be a better option?

**Response**: The PPV was used primarily to identify detected plumes with a large number of pixels that were incorrectly assigned to the source. To measure the overall performance, the PPV needs to be used complementary with the TP. A "good" detection would thus need to have both high TP and PPV. The detection in the example has a high PPV but also a low TP.

We did not use the Hanssen-Kuiper Skill Score / True Skill Statistics, because it depends on the number of true negatives (TN), i.e. all pixels outside the true plume. TN cannot be defined unambiguously in our application, because it depends on swath width and the size of the model domain.

Please note that in the paper we do not evaluate the performance of the plume detection algorithm but the performance of the different instrument scenarios for detecting the $CO_2$ plume. We have revised Section 2.2.2 ("Performance evaluation") to better explain this purpose (see also our response to Comment 16).

**6. Reviewer #1**: To my knowledge, the Z-test is appropriate only for samples with more than 30 elements. 5 out of 8 tested smoothing kernels (Fig.4) consist of less elements. Why do you think the z-test is still appropriate in these cases?

**Response**: The Z-test can also be applied to smaller samples, even to single observations, when the variance of the population is known. The sample size is only a problem if the variance has to be estimated from the sample, which is not the case for our application.

**7. Reviewer #1**: What happens at cloud edges and at the edges of the swath? Effectively, the smoothing kernel here has fewer elements. Is this considered?

**Response**: Yes, the n in Eq. (1) is smaller if satellite pixels are missing in the neighborhood. We modified and added the following sentences to explain the difference between n and ns: >>> *The random component would reduce with the inverse square root of the number of valid pixels n in the neighborhood, whereas the systematic error would be approximately independent of n. underline{The number of valid pixels*

**Specific comments:**

**8. Reviewer #1**: P2L23: Please also cite Reuter et al. 2014 in this context. They analyzed CO2+NO2 measurements from the same sensor (SCIAMACHY) and proposed multi-species measurements for future satellite missions. As signals in SCIAMACHY data are smaller because of the large pixel size, they had to follow a statistical approach. Note that this does not only cover strong localized sources, but all localized sources.

**Response**: We added the reference.

**9. Reviewer #1**: P3L4-L5: I would suggest moving this outlook to the conclusions section.

**Response**: We think it is better to keep this here to describe the context within the SMARTCARB project.

**10. Reviewer #1:** P3L8-L27: Strictly speaking, everything in these lines does neither belong to "data" nor to methods. Maybe it would fit better in the introduction.

**Response**: We moved the section to the introduction.

**11. Reviewer #1**: P3L24: The situation is also not that simple for individual point sources, because chemistry changes the NO2 plume which is not the case for the CO2 plume. The following points should be addressed in this context: 1) NO2 decays with time. 2) NO2 and CO2 can be emitted/transported into various altitudes with differences chemical regimes. For example, the high altitude part of the plume may decay faster/slower compared with the low altitude part of the plume. As the wind direction often changes with height, the satellite observed CO2 plume may have a different (wider) shape as the NO2 plume. 3) NO2 is primarily emitted as NO (which later forms NO2), which can result in different plume shapes especially near the source.

**Response**: Our simulations account for (1) the NO2 decay with time with a constant exponential decay and (2) the conversion of NO to NO2 by emitting and transporting NO2 as part of a NOx tracer. The last paragraph of the introduction has been changed as follows:

>>> *However, this requires that the plumes of CO2 and of the auxiliary trace gas are spatially congruent. This is usually the case when they are emitted from the same source, for example, a power plant. The shape of the NO2 plumes might deviate from the CO2 plume for two reasons. First, NO2 is emitted mainly as nitrogen monoxide (NO), which is converted to NO2 over time resulting in lower NO2 concentrations near the source. Second, NO2 decays slowly with time reducing NO2 concentrations downstream. To account for these two effects, we simulated nitrogen oxides (NOX = NO + NO2) that slowly decays with time and calculated NO2 from NOX concentrations offline by applying a formula frequently used to represent NO2:NOX ratios downstream of emission sources (Düring et al. 2011). The situation is more complex for cities where the emissions originate from different sectors (industry, heating, transport*

*etc.) that emit with different temporal profiles and at different altitude levels, and which have different emission ratios (Brunner et al. 2019). In this study, we therefore carefully consider the vertical and temporal profiles of emissions from different sectors, which makes it possible to test for congruence.* <<<

*We also added the following sentence to Section 2.1.1*

>>> *NO2 concentrations were calculated from NOX concentrations using an empirical formula used frequently for representing NO2:NOx ratios downstream of emission sources (Düring et al. 2011).* <<<

and the following sentence to the abstract:

>>> CO2 and NO2 plumes were found to overlap to a large extent, although NOX had a limited lifetime (assumed to be 4 hours) and although CO2 and NOX were emitted with different NOX:CO2 emission ratios by different source types with different temporal and vertical emission profiles. <<<

The effect of different chemical regimes on the plumes shapes might be important, but we are not aware of any publication studying the effect or showing that it has a significant effect on the congruence of CO2 and NO2 plumes. Our study is also not able to quantify this effect, because it requires full chemistry transport simulations, which is computational too expensive for one year of simulation. Because we think that the introduction is not a good place to discuss this, we added the following sentence to the discussions (Section 4.2) instead:

>>> *Our study could not simulate detailed NOx chemistry due to the high computational costs. We can therefore not rule out a larger mismatch between some CO2 and NO2 plumes than simulated here, for example, because of different NO2 decay rates at different altitudes. However, Reuter et al. (2019) found no obvious mismatch when comparing co-located CO2 and NO2 plumes from OCO-2 and TROPOMI observations. A follow-up study for quantifying the effect of NOx chemistry would be certainly desirable.* <<<


**12. Reviewer #1**: P7L4: sun-synchronous orbits are not circular.

**13. Reviewer #1**: P7L5: "… in case of 3 satellites." Please describe what this means (the satellites will have the same Eq. crossing times…).

**Response**: We re-formulated the paragraph for clarification:

>>> The basic assumption for a constellation is that the individual satellites are spaced with equal angular distance in the same orbit with the same orbit parameters, for instance separated by 180°, 120° and 90° on a full circle in case of 2, 3 and 4 satellites. <<<


**14. Reviewer #1**: P9L30: Cloud coverage is a main driver for data yield of XCO2 satellite products. I guess COSMO diagnoses fractional cloud coverage in its 3D model domain. How do you/COSMO compute the 2D cloud fractional coverage from this (e.g., maximum/minimum overlap assumption) and how realistic are the 2D cloud coverage statistics for this particular model?

**Response**: We'd like to thank reviewer #1 for this important comment. Cloud cover is computed by COSMO assuming minimum overlap. We investigated how cloud cover in COSMO compares with MODIS satellite data and added the following sentences to the method section proving more details on COSMO cloud fraction and the reason why it was used in our study:

Furthermore, we added a paragraph to Section 4.1 comparing the difference in data yield using COSMO cloud fractions and the MODIS cloud mask product.

We will add the following figures to the supplement of the paper:

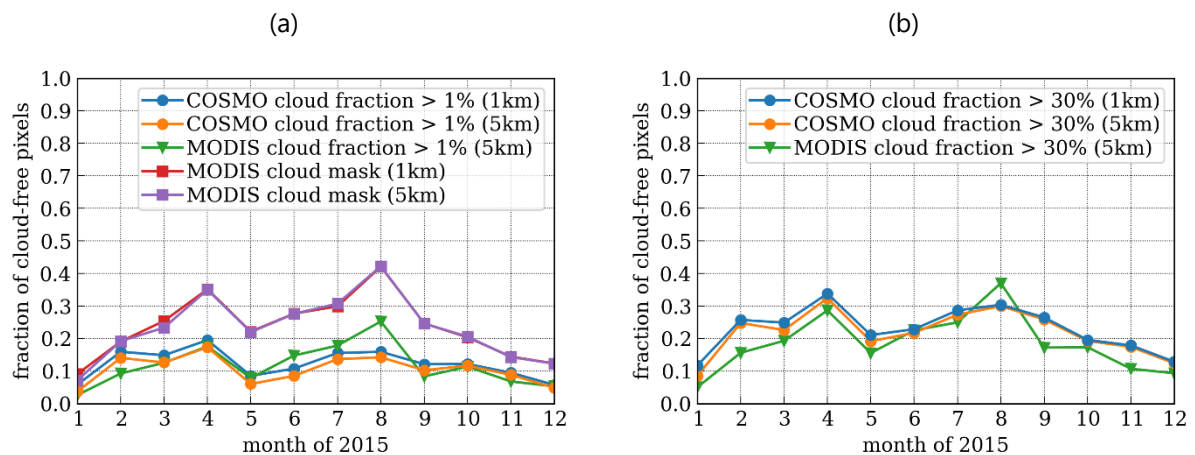(a)                                                            (b)



*Figure S1: Monthly fractions of cloud-free pixels using COSMO- and MODIS-based cloud masks with different spatial resolution (1 km and 5 km at nadir) and (a) 1% and (b) 30% cloud fraction as threshold.*

The conclusions were also updated to include the new result:

"*This number is somewhat smaller than reported in earlier studies (Bacour et al. 2015, Pillai et al., 2016), mainly because masking cloudy pixels based on the simulated cloud fields leads to less cloud-free observations than using the MODIS cloud mask product.*"

**15. Reviewer #1**: P10 Fig3 Caption: "… are marked with "T" for True." Shouldn't be True = blue/red and

T = detected by Z-test? If so, consider replacing T by Z.

**Response**: "T" was used to mark pixels detected by the Z-test. We changed "T" to "D" to avoid confusion with the true plume and changed caption and legend accordingly.

**16. Reviewer #1**: P11L7: The choice of neighborhood size appears a bit arbitrary. Please discuss, why it makes sense to try neighborhood sizes as selected. For example, I could imagine, that the maximum meaningful neighborhood size is reached once the noise of the smoothed pixels dropped significantly below levels of typical enhancements. On the other hand, the neighborhood should probably never be significantly larger than the typical plume extends.

**Response**: The performance of different neighborhoods was analyzed in more detail in the SMART-CARB final report (see https://www.empa.ch/documents/56101/617885/FR_Smartcarb_final_Jan2019.pdf), which is also cited in the manuscript. We added the following sentences to the 2$^{nd}$ paragraph of Section 3.2 ("Signal-to-noise ratios") to provide some more details:

>>> A large neighborhood reduces the random noise of the measurements and therefore allows detecting smaller signals. On the other hand, a large neighborhood will include background values in the computation of the averages at the plume edges and reduce the signal. The sizes used here roughly correspond to the typical diameters of the CO2 plumes from Berlin (about 15 km) and the power stations (about 6 km), respectively, and were also found most suitable for the plume detection algorithm, because they maximize TPs without reducing PPVs too much (see also Kuhlmann et al. 2019). <<<

**17. Reviewer #1**: P11L20,L21: Please describe why the Z-test calculates the SNR (e.g., the variance of the basic population is assumed to be known from the measurement errors, Xobs is an estimator for the expectation value).

**Response**: We added a more detailed description of the z-test and it is relation to the signal-to-noise ratio (see revised manuscript).

**18. Reviewer #1**: P11Eq1: On P9L4 you stated that systematic errors are not scope of this study. Why considering them in Eq1?

**Response**: The equation (Eq. 2 in the revised manuscript) provides a general form that considers both random and systematic errors. In our study, we do not consider systematic errors in the satellite observations but we still consider systematic errors in the background field resulting from the biosphere or other anthropogenic sources.

**19. Reviewer #1:** Xobs are the smoothed observations (not to be confused with the original observations). This should be corrected also at P11L13.

**Response**: We changed this to "spatially averaged satellite observations"


**20. Reviewer #1:** Sigma_rand is only approximately the same for all soundings within the smoothing kernel but strictly speaking; it is different for each sounding.

**Response**: This is correct. We used a constant value, because the random error in real data is also not known exactly.


**21. Reviewer #1**: P11Eq2: Does this imply that there will also be false positives just by chance? How often is this expected to happen?

**Response**: Yes, since we used q=0.99, we expect that 1% of detected pixels are false positives. However, the number of false positives is reduced by labelling in the second step of the algorithm.


**22. Reviewer #1**: P11L30: Does this mean $sigma_{sys}$ = $sigma_{bg}$ in Eq 1?

**Response**: Yes, under the assumption made in our study.


**23. Reviewer #1**: P12L1: What is the chosen level for significance?

**Response**: In this study, we used q=0.99.


**24. Reviewer #1**: P12L5: Please explain "Moor neighborhood".

**Response**: In (digital) image processing, a "Moore neighborhood" is a neighborhood where each pixel has eight potential neighbors (see for example https://en.wikipedia.org/wiki/Pixel_connectivity). We restructured the paragraph slightly to improve readability:

>>> *In the second step, pixels that are enhanced (``true'') and connected are assumed to belong to the same plume. We label regions of connected pixels using a standard labeling algorithm. Neighboring pixels are identified using a Moore neighborhood, where each pixel has eight potential neighbors. Each region is assigned a unique integer value.* <<<


**25. Reviewer #1**: P12L9: "… radius of 5km" In reality, this can result in non-detected NO2 plumes when NO2 is emitted primarily as NO.

**Response:** Yes this is correct and we accounted for this by calculating NO2 columns from NOx fields with an empirical formula (see our reply to comment #11). It should be noted that we did not find a scene where a radius of 5 km was not sufficient for assigning a detected (cloud-free) plume to the source.

**26. Reviewer #1**: P12L21: Can you separate both effects; which one usually dominates? In other words, why is a plume not detected, too low signal or CO2 plume != NO2 plume?

**Response**: The NO2-based plume detection often fails, if the plume overlaps with neighboring plumes which results in low PPV or if the scene is extremely cloudy. The NO2 signal-to-noise ratio is generally high and the CO2 and NO2 plume match well in the simulations.

**27. Reviewer #1**: P13Fig.5d: How can you compute Xobs in totally cloudy regimes?

**Response**: $X_{obs}$ cannot be computed for fully cloudy scenes. Figure 5 shows examples of true CO2 plumes taken from the synthetic observations. We modified the caption for clarification.

**28. Reviewer #1**: P14L14: If the standard deviation of the background was used as threshold, shouldn't there be about 16

**Response**: Sorry, but the question seems incomplete.

**29. Reviewer #1**: P26 Tab8: What does "large number of false positives" mean. How large is large?

**Response**: A computation of the number of false positives was not possible due to the missing true plume for Janschwalde. Instead, visual inspection was used, but "large number of false positives" looks typically like Figure 13c. We now cite Fig. 13c as example both in the caption as well as in the text.

**30. Reviewer**: P27L29: Why should the higher emissions and the not-considering of emission profiles by Pillai et al., result in a larger number of usable plumes? To my knowledge, they analyzed each sufficiently cloud free overpass; please check. Isn't it expected, that you find less usable scenes simply because inverse modeling studies usually assume knowledge of the plume position while you have to search for strong enough signals?

**Response**: In this section, we compare potentially observable CO2 plumes independent of the detection algorithm and therefore expected that our numbers would be similar to potentially useful plumes identified by Pillai et al. (2016). After checking Pillai et al (2016), we decided that it might be more reasonable to compare their number of potentially useful orbits (41 for 500-km swath) instead of the number of orbits with successful emission estimates (27 for 500-km swath). We therefore revised the paragraph as follows:

>>> *For Berlin, we estimated that 3 to 13 potentially useful CO2 plumes (defined as plumes with at least 100 cloud-free pixels above a threshold of 0.05 ppm) would be observable, but not necessarily detectable, during one year by a single CO2M satellite with a 250-km wide swath. Pillai et al. (2016) identified 41 potentially useful orbits for estimating emissions with a 500-km wide swath. Although a direct comparison of these two numbers is difficult because of the different swath widths and the different definitions of "usefulness", we can still conclude that our study identified significantly fewer plumes than Pillai et al. (2016) even after halving their number to account for their wider swath.* <<<

Next, we add new paragraph showing that the difference can be mostly be explained by differences in the cloud masks used in our and their study (see response to Comment 14).

Higher emissions and not-considering emission profiles result in more potentially useful plumes, because they result in larger plumes that are less likely fully covered in clouds. We revised the paragraph as follows:

>>> *Further differences are to be expected because small plumes with less than 100 pixels were excluded in our case and because of different meteorology, especially cloud cover and wind speed, for the different simulation periods. In addition, Pillai et al. (2016) used higher emissions and did not consider vertical profiles of emissions, which results in stronger and correspondingly larger plumes, which are less likely to be fully covered by clouds (Brunner et al., 2019).* <<<

**31. Reviewer #1**: P28L4: ...and also due to the fact that NO2 is used primarily for plume detection but XCO2 will also be used for flux estimation.

**Response**: Sorry, we are not able to relate the comment with our manuscript.

**32. Reviewer #1**: P30L7,L8: Which effect dominates, the higher SNR or the lower sensitivity to clouds (see general comment)?

**Response**: (see our reply to Comment 3)

**33. Reviewer #1**: P30L24,L25: ... and cleaner sources in terms of NO2 emissions.

**Response**: The sentence now reads: "Therefore, adding an NO2 instrument is highly recommended and the low-noise instrument is preferable for detecting also weaker and cleaner plumes in terms of NO2 emissions."

**34. Reviewer #1**: P30L26: Please add a discussion that this future plume detection scheme will have to solve the problem of the unknown BG before it can be applied to real data.

**Response**: (see our reply to Comment 1)

**Technical corrections**

**Reviewer #1**: P3L1: can be detected -> is expected to be detected

**Response**: Done

**Reviewer #1**: P3L1: CO2 (or NO2 or CO) -> CO2, NO2, or CO

**Response**: Done.

**Reviewer #1**: P7L2: Please add Eq. crossing times to Tab3.

**Response:** Equator crossing times are already in Table 2.

**Reviewer #1**: P8L3: "...requirements for Sentinel-5". Please add a reference if possible.

**Response**: We added the following reference:

Ingmann et al (2012): Requirements for the GMES Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p, doi: 10.1016/j.rse.2012.01.023.


**Reviewer #1**: P10 Fig3 Caption: "(a) A large (red) and small (blue) …" Have you mixed up large and small?

**Response**: Yes, the caption is correct now.


**Reviewer #1**: P11L7: "at a satellite" -> "at a smoothed satellite"

**Response**: Changed to: "at a spatially smoothed satellite pixel."


**Reviewer #1**: P11L12+L13: "n" -> "ns"

**Response**: We modified the sentence (see our previous answer).


**Reviewer #1**: P27L15: "which we describe in another publication" do you mean "which we will describe in another publication"

**Response**:  Changed.


**Reviewer #1**: P27L17: "used our study" -> "used in our study"

**Response**: Changed.


**Reviewer #1**: P27L29: remove "using"

**Response**: Removed

# Detectability of $CO_2$ emission plumes of cities and power plants with the Copernicus Anthropogenic $CO_2$ Monitoring (CO2M) mission

Gerrit Kuhlmann[1], Grégoire Broquet[2], Julia Marshall[3], Valentin Clément[4,5], Armin Löscher[6], Yasjka Meijer[6], and Dominik Brunner[1]

[1]Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland
[2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[3]Max Planck Institute for Biogeochemistry (MPI-BGC), Jena, Germany
[4]Center for Climate Systems Modelling (C2SM), ETH Zurich, Zurich, Switzerland
[5]MeteoSwiss, Kloten, Switzerland
[6]European Space Agency (ESA), ESTEC, Noordwijk, The Netherlands

**Correspondence:** Gerrit Kuhlmann (gerrit.kuhlmann@empa.ch)

**Abstract.** High-resolution atmospheric transport simulations were used to investigate the potential for detecting carbon dioxide ($CO_2$) plumes of the city of Berlin and neighboring power stations with the Copernicus Anthropogenic Carbon Dioxide Monitoring (CO2M) mission, which is a proposed constellation of $CO_2$ satellites with imaging capabilities. The potential for detecting plumes was studied for satellite images of $CO_2$ alone or in combination with images of nitrogen dioxide ($NO_2$) and
5 carbon monoxide (CO) to investigate the added value of measurements of other gases co-emitted with $CO_2$ that have better signal-to-noise ratios. The additional $NO_2$ and CO images were either generated for instruments on the same CO2M satellites ($2\times2$ km$^2$ resolution) or for the Sentinel-5 instrument ($7\times7$ km$^2$) assumed to fly two hours earlier than CO2M. Realistic $CO_2$, CO and $NO_X$ (= NO + $NO_2$) fields were simulated at $1\times1$ km$^2$ horizontal resolution with the Consortium for Small-scale Modeling - Greenhouse Gases (COSMO-GHG) model for the year 2015, and used as input for an orbit simulator to generate
10 synthetic observations of columns of $CO_2$, CO and $NO_2$ for constellations of up to six satellites. A ~~new~~ simple plume detection algorithm was applied to detect coherent structures in the images of $CO_2$, $NO_2$ or CO against instrument noise and variability in background levels. Although six satellites with an assumed swath of 250 km were sufficient to overpass Berlin on a daily basis, only about 50 out of 365 plumes per year could be observed in conditions suitable for emission estimation due to frequent cloud cover. With the $CO_2$ instrument only 6 and 16 of these 50 plumes could be detected assuming a high ($\sigma_{VEG50}$
15 = 1.0 ppm) and low noise ($\sigma_{VEG50}$ = 0.5 ppm) scenario, respectively, because the $CO_2$ signals were often too weak. A CO instrument with specifications similar to the Sentinel-5 mission performed worse than the $CO_2$ instrument, while the number of detectable plumes could be significantly increased to about 35 plumes with an $NO_2$ instrument. $CO_2$ and $NO_2$ plumes were found to overlap to a large extent, although $NO_X$ had a limited lifetime (assumed to be 4 hours) and although $CO_2$ and $NO_X$ were emitted with different $NO_X$:$CO_2$ emission ratios by different source types with different temporal and vertical emission
20 profiles. Using $NO_2$ observations from the Sentinel-5 platform instead resulted in a significant spatial mismatch between $NO_2$ and $CO_2$ plumes due to the two hours time difference between Sentinel-5 and CO2M. The plumes of the coal-fired power plant Jänschwalde were easier to detect with the $CO_2$ instrument (about 40-45 plumes per year), but again, an $NO_2$ instru-

**1**

ment ~~performed significantly better~~ could detect significantly more plumes (about 70~~plumes~~). Auxiliary measurements of $NO_2$ were thus found to greatly enhance the capability of detecting the location of $CO_2$ plumes, which will be invaluable for the quantification of $CO_2$ emissions from large point sources.

## 1 Introduction

5   The signatory countries of the Paris climate agreement have set ambitious goals to reduce $CO_2$ emissions and limit global warming to below 2° C above pre-industrial levels (UNFCCC, 2015). The efficient implementation and management of long-term policies will require consistent, reliable, and timely information on $CO_2$ emissions (Ciais et al., 2015; Pinty et al., 2018). The majority of these emissions are concentrated on a small fraction of the globe, primarily on cities and power plants. Acknowledging their important role, cities have started to devise policies for cutting $CO_2$ emissions often surpassing the reduction

10  targets of the respective countries (e.g. C40 cities, 2018). However, many cities are currently lacking ~~an independent~~ a $CO_2$ emission monitoring system to evaluate their policies.

     The European Space Agency (ESA) and the European Commission (EC) therefore propose the Copernicus Anthropogenic Carbon Dioxide Monitoring (CO2M) mission, a constellation of $CO_2$ satellites with imaging capability, to support the quantification of anthropogenic $CO_2$ fluxes and to assist greenhouse gas mitigation policies at the national, city and facility level

15  (Sierk et al., 2019). The satellites are envisioned as an essential component of a $CO_2$ observation system to be established under Europe's Earth observation program Copernicus (Ciais et al., 2015; Pinty et al., 2018). The system would allow for observing $CO_2$ plumes of individual point sources such as large cities and power plants and for quantifying the respective emissions during single satellite overpasses (Bovensmann et al., 2010; Pillai et al., 2016; Velazco et al., 2011). A $CO_2$ plume is defined here as an enhancement of $CO_2$ concentrations above the background in the satellite image caused by the emissions

20  of a given source. The emissions of the source can be estimated from the $CO_2$ enhancement inside the plume, which requires that the plume location is identified in the satellite observations. An atmospheric transport model may be used for simulating the plume location and for estimating the emissions with an inversion framework (e.g. Pillai et al., 2016; Broquet et al., 2018). However, the simulated plume might be significantly displaced due to uncertainties in wind fields and emission heights, which would result in systematic errors in the estimated emissions (Broquet et al., 2018; Brunner et al., 2019). It is therefore desirable

25  to detect the plume directly in the satellite observations, which would make it possible to correct transport-related errors in the simulations but also to estimate the emissions directly from the $CO_2$ enhancements in the plume using plume fitting or mass balance approaches, which only require an estimate of the mean wind speed within the plume (Fioletov et al., 2015; Krings et al., 2013; Varon et al., 2018). While the potential for detecting and estimating emissions from $CO_2$ fluxes has been demonstrated for strong $CO_2$ plumes of megacities and large point sources ~~has been demonstrated for~~

30  using the Orbiting Carbon Observatory 2 (OCO-2, Crisp et al. (2017)) (Nassar et al., 2017; Reuter et al., 2019), a major challenge is to accurately determine the location of weaker $CO_2$ plumes with signal-to-noise ratios near or below the detection limit for single pixels. The detection of such weak plumes is additionally challenged by the interference with signals from biospheric $CO_2$ fluxes and other anthropogenic sources in the vicinity of the target. Therefore, measurements

of auxiliary trace gases co-emitted with $CO_2$ but little affected by biospheric processes such as carbon monoxide (CO) and nitrogen dioxide ($NO_2$) were proposed to help separate anthropogenic from biospheric $CO_2$ signals ~~(Ciais et al., 2015)~~ (Reuter et al., 2014; Ciais et al., 2015).

This study presents results from the SMARTCARB project (use of Satellite Measurements of Auxiliary Reactive Trace gases for fossil fuel CARBon dioxide emission estimation (Kuhlmann et al., 2019)), which aimed to assess the potential synergies of measurements of CO and $NO_2$ for observing and quantifying $CO_2$ emissions and to help define the required satellite specifications. To address these questions, Observing System Simulation Experiments (OSSE) were conducted, for which synthetic satellite observations were generated from high-resolution atmospheric transport simulations. The model domain was centered on the city of Berlin and also covered several nearby power plants. Similar simulations were already performed in previous OSSEs (Pillai et al., 2016; Broquet et al., 2018), but they did not have a comparable spatial resolution or temporal extent or did not cover the additional species $NO_2$ and CO as investigated here.

In a companion paper Brunner et al. (2019) presented the overall model setup and emphasized the importance of properly accounting for the vertical placement of $CO_2$ emissions from large point sources in atmospheric $CO_2$ simulations. Here, we investigate whether and how often the $CO_2$ ~~(or~~, $NO_2$ or CO ~~)~~ plume of a city or a power plant ~~can~~ is expected to be detected during a year depending on the size of the satellite constellation and on instrument specifications. For this purpose, a ~~novel~~ simple plume detection algorithm was developed which identifies the plume signals against instrument noise and background variability. In a follow-up study, we will quantify the emissions from Berlin and a few power plants in the model domain from the synthetic satellite observations using both inverse and mass-balance approaches, building on the plume detection presented here.

## 2 ~~Data and methods~~

In a satellite image, a plume may be defined as a collection of spatially connected pixels with elevated signals starting at a source. Whether and how frequently the plume of a given source can be detected depends on several, partly interdependent factors:

- The number of satellites and the instrument's swath width, as they determine the number of overpasses over the plume and how much of the plume is visible in the satellite image.

- The intensity of the emission source, which affects the amplitude of the enhancement above background.

- The meteorological conditions, notably wind speed and turbulence, which determine the dilution and dispersion of the emissions.

- The single sounding precision of the instrument, which determines if the enhancement within the plume can be detected.

- The variability of the background, which is caused by anthropogenic emissions and biospheric fluxes ~~upstream or downstream of the considered~~ in the vicinity of the source and which is additionally affected by meteorology.

– The presence of clouds partially or fully obscuring the plume.

Since most of these factors vary with season, the detectability also depends on the time of the year. Therefore, long simulations covering a full year were conducted.

Because the detection of weak anthropogenic $CO_2$ plumes is affected by interference with biospheric $CO_2$ signals, auxiliary trace gases co-emitted with $CO_2$ could be used for locating the $CO_2$ plume in the satellite image. However, this requires that the plumes of $CO_2$ and of the auxiliary trace gas are spatially congruent. This is usually the case when they are emitted from the same source, for example, a power plant. The shape of the $NO_2$ plumes might deviate from the $CO_2$ plume for two reasons. First, $NO_2$ is emitted mainly as nitrogen monoxide (NO), which is converted to $NO_2$ over time resulting in lower $NO_2$ concentrations near the source. Second, $NO_2$ decays slowly with time reducing $NO_2$ concentrations downstream. To account for these two effects, we simulated nitrogen oxides ($NO_X = NO + NO_2$) that slowly decays with time and calculated $NO_2$ from $NO_X$ concentrations offline by applying a formula frequently used to represent $NO_2:NO_X$ ratios downstream of emission sources (Düring et al., 2011). The situation is more complex for cities where the emissions originate from different sectors (industry, heating, transport etc.) that emit with different temporal profiles and at different altitude levels~~and~~, and which have different emission ratios ~~of to or CO~~ (Brunner et al., 2019). In this study, we therefore carefully consider the vertical ~~distribution of emissions for different species~~and temporal profiles of emissions from different sectors, which makes it possible to test for congruence.

## 2 Data and methods

### 2.1 Synthetic satellite observations

#### 2.1.1 Model simulations

The synthetic satellite observations were generated from high-resolution simulations conducted with the COSMO-GHG model. COSMO is a hydrostatic, ~~limit-area~~ limited-area model developed by the Consortium for Small-scale Modeling (Baldauf et al., 2011), for which an extension has been developed for the simulation of greenhouse gases (COSMO-GHG)(Oney et al., 2015; Liu et al., 2017).

COSMO-GHG was set up to simulate $CO_2$, CO and $NO_x$ concentration fields for nearly the complete year 2015 (1 January - 25 December). The model domain extended about 750 km in east-west and 650 km in south-north direction. It was centered over the city of Berlin and also covered numerous power plants in Germany and neighboring countries. The spatial resolution was 1.1 km × 1.1 km horizontally with 60 vertical levels up to an altitude of 24 km. Figure 1 presents the model domain and marks the location of Berlin and the six largest coal-fired power plants. The detailed model setup is described by Brunner et al. (2019).

Initial and lateral boundary conditions (ICBC) for meteorological variables were provided by the operational European COSMO-7 analyses of MeteoSwiss with hourly temporal and 7 km horizontal resolution. For the tracers, ICBCs were obtained from the European Center for Medium Range Weather Forecast (ECMWF) through the European Earth observation program
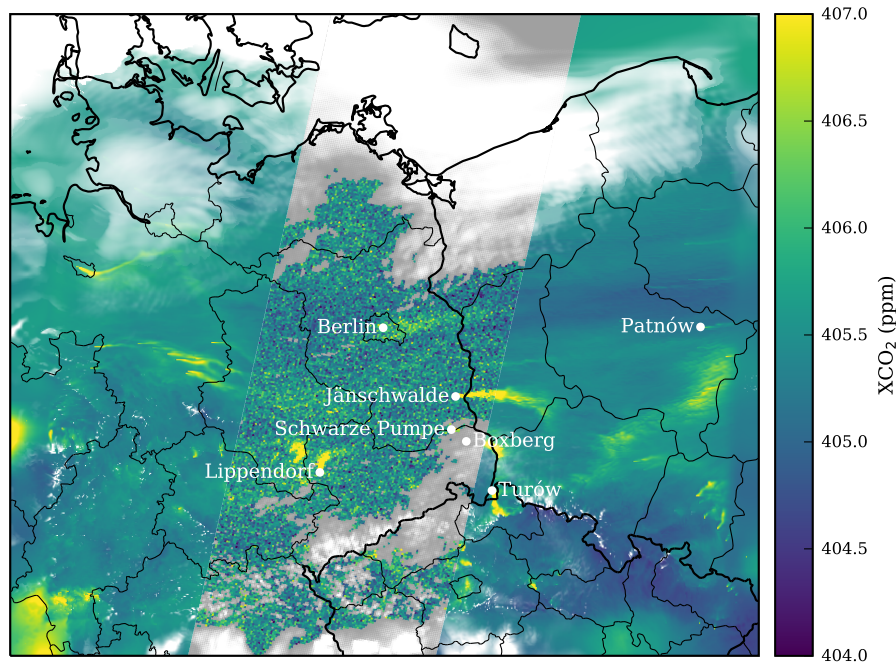
**Figure 1.** Simulated $XCO_2$ field on 23 April 2015 in the SMARTCARB model domain overlaid with an example of a 250 km wide swath of the planned Sentinel $CO_2$ instrument (low noise scenario). Missing $CO_2$ measurements are shown in gray. Cloud cover is overlayed in white with transparency corresponding to total cloud fraction.

**Table 1.** Emissions of largest power plants in model domain according to TNO/MACC-3 inventory for the year 2011 as used in this study.

| Power plant | $CO_2$ (Mt yr$^{-1}$) | $NO_x$ (kt yr$^{-1}$) | CO (kt yr$^{-1}$) |
|---|---|---|---|
| Jänschwalde | 33.3 | 26.9 | 44.1 |
| Boxberg | 19.0 | 15.4 | 14.4 |
| Lippendorf | 15.3 | 12.3 | 2.7 |
| Turów | 8.7 | 13.1 | 1.3 |
| Schwarze Pumpe | 8.2 | 6.6 | 5.3 |

Copernicus. $CO_2$ and CO boundary conditions were taken from a global free-running $CO_2$ simulation with 137 levels and about 15 km horizontal resolution (T1279 spectral resolution, experiment "gf39", class "rd")(Agustí-Panareda et al., 2014). For NO and $NO_2$, boundary conditions were taken from ECMWF's operational global forecasts for aerosol and chemical species with 60 vertical levels and a horizontal resolution of about 60 km (T255 resolution, experiment "0001", class "mc") (Flemming et al., 2015).

Anthropogenic emissions were obtained by combining the Netherlands Organisation for Applied Scientific Research European Monitoring Atmospheric Composition and Climate version 3 (TNO/MACC-3) inventory (Kuenen et al., 2014, for Version 2) with a detailed inventory provided by the city of Berlin (AVISO GmbH and IE Leipzig, 2016). The temporal variability of emissions was accounted for by applying diurnal, weekly, and seasonal cycles according to Selected Nomenclature for Air Pol-

lution (SNAP) categories. Furthermore, emissions were vertically distributed using specific vertical profiles for the different emissions categories and plume rise calculations for six largest power plants and the major point sources in Berlin (Brunner et al., 2019). Hourly biospheric fluxes both photosynthesis and respiration were generated with the vegetation photosynthesis and respiration model (VPRM) at the resolution of the COSMO model (Mahadevan et al., 2008).

5    According to the official inventory of the city of Berlin, total annual $CO_2$ emissions of Berlin were $16.9\,\mathrm{Mt\,CO_2\,yr^{-1}}$ (in the reference year 2012 of the inventory). This is about a factor two smaller than in previous studies, e.g. in the LOGOFLUX project (Chimot et al., 2013; Bacour et al., 2015; Pillai et al., 2016), which relied on unrealistically high emissions as provided by the global EDGAR inventory (Version 4.1). Due to the diurnal cycle of emissions, emissions were somewhat larger (about $20.0\,\mathrm{Mt\,CO_2\,yr^{-1}}$) around the time of the satellite overpasses (10-11 UTC). Table 1 summarizes the $CO_2$, $NO_x$ and CO
10  emissions of the five largest power plants in the domain.

The simulations included a total of 50 different passively transported tracers representing the three different gases further divided into different sources, release times or release altitudes. This also included background tracers constrained at the lateral boundaries by the global-scale models and two tracers for biospheric respiration and photosynthesis for $CO_2$. Due to the reactivity of $NO_x$, five different $NO_x$ tracers with e-folding lifetimes of 2, 4, 12 and 24 hours and infinity were included,
15  considering that the lifetime of $NO_x$ varies between about 2 and 24 hours (Schaub et al., 2007). The full list of tracers is provided in the SMARTCARB final report (Kuhlmann et al., 2019, p. 15f)

In this study we use the following seven tracers that have been computed from the 50 tracers in the simulations:

- X_BER: concentrations from time-varying emissions of Berlin,

- X_PP: concentrations from time-varying emissions from the six largest power plants in the model domain,

20
- X_ANTH: concentrations from other anthropogenic sources in the domain excluding emissions of Berlin (X_BER) and the six largest power plants (X_PP),

- X_BIO: concentrations from local biospheric fluxes, i.e. respiration and photosynthesis within the domain (only for $CO_2$),

- X_TOT: concentrations from all emissions and biospheric fluxes as well as inflow from lateral boundaries,

25
- X_BER_BG: concentrations from emissions, fluxes and lateral boundaries excluding emissions from Berlin (= X_TOT – X_BER),

- X_PP_BG: concentrations from emissions, fluxes and lateral boundaries excluding emissions from the six major power plants (= X_TOT – X_PP),

where X is $CO_2$, CO or $NO_2$. $NO_2$ concentrations were calculated from $NO_x$ concentrations using an empirical formula
30  used frequently for representing $NO_2$:$NO_x$ ratios downstream of emission sources (Düring et al., 2011). For $NO_2$ only the tracers with a lifetime of 4 hours were used. Note that only the sum of the emissions from the six power plants was simulated

**Table 2.** Satellite platforms and orbits.

| Parameter | CO2M satellite | MetOp-SG-A |
|---|---|---|
| Orbit type | Sun-synchronous | Sun-synchronous |
| Inclination | 97.77° | 98.7° |
| Orbits per day | 14+10/11 | 14+6/29 |
| Cycle duration | 11 days | 29 days |
| Cycle length | 164 orbits | 412 orbits |
| Altitude | 602.24 km | 830.16 km |
| Orbit Period | 96.58 minutes | 101.36 minutes |
| Local time in descending node (equator crossing time) | 11:30 hrs | 9:30 hrs |

**Table 3.** Observation geometries for instruments on the CO2M satellite and Sentinel-5.

| Parameter | CO2M satellite | Sentinel-5 |
|---|---|---|
| Number of across-track pixels | 125 | 208 |
| Swath | 250 km | 2670 km |
| Field of view | 23.22° | 107.1° |
| Pixel size | $2 \times 2$ km$^2$ | from $7.5 \times 7.5$ km$^2$ (nadir) to $35 \times 7.5$ km$^2$ (swath edge) |
| Along-track sampling time | 0.286 seconds | 1.13 seconds |

but not the power plants individually, which often complicated the analysis due to overlapping plumes. For the analysis, the three-dimensional model fields were vertically integrated to compute column-averaged dry air mole fractions of $CO_2$ (XCO$_2$). Likewise, tropospheric CO and $NO_2$ vertical column densities (VCD) were generated by considering only the model fields below 10 km altitude.

### 5 2.1.2 Satellite instrument scenarios

For the $CO_2$, CO and $NO_2$ satellite observations, different instrument scenarios were prescribed by ESA for the study in terms of orbit, spatial resolution and spatial and temporal coverage. For the scenarios, two representative platforms were considered: MetOp-SG-A (Meteorological Operational Satellite – Second Generation - A) and the proposed CO2M satellite. MetOp-SG-A is a series of meteorological satellites in low earth orbit that will carry the Sentinel-5 instrument measuring, among others, $NO_2$
10 and CO VCDs. The CO2M mission is a proposed platform with a $CO_2$ instrument as the main payload and optionally with instruments for $NO_2$, CO and aerosols. Both satellites will be in sun-synchronous orbits but with different equator crossing times and repeat cycles.

In addition to a single satellite, the potential of a constellation of multiple CO2M satellites was also studied. The basic assumption for a constellation is that the individual satellites ~~would be spaced equally in a common (circular) orbit~~ are spaced
15 with equal angular distance in the same orbit with the same orbit parameters, for instance separated by 180°, 120° and 90° on a full circle in case of ~~3 satellites. Here, we analyze constellations between one and six~~ 2, 3 and 4 satellites. The individual
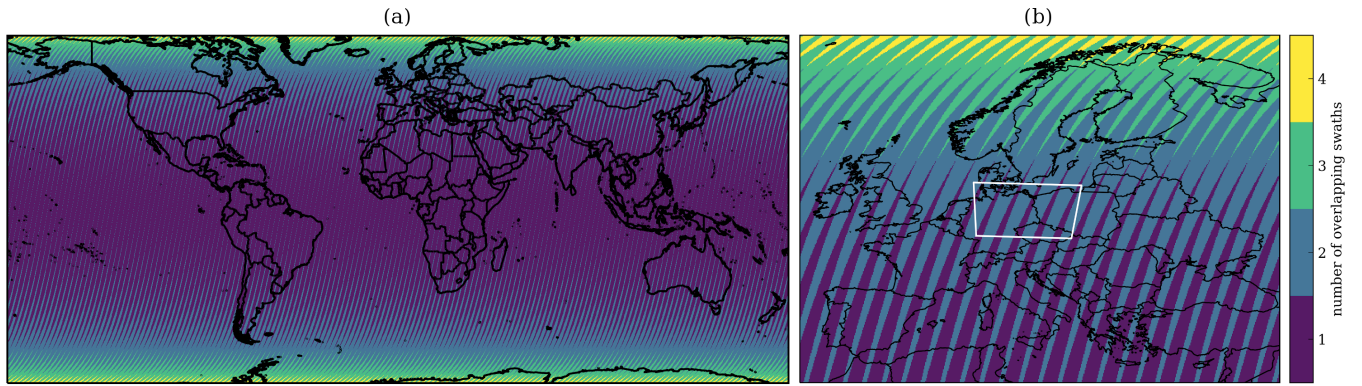
7

**Figure 2.** Spatial coverage of one CO2M satellite within its 11-day repeat cycle (a) globally and (b) over Europe. The white square marks the COSMO-GHG model domain in which the number of overpasses is either one or two. The exact locations of the "stripes" are arbitrary and depend on the equator starting longitude (here: 0 °E) of the satellite.

satellites can be distinguished by their starting longitude at the equator of the first orbit in the repeat cycle. Here, we analyze constellations between one and six satellites.

For the computation of orbits, we adopted the orbit simulator of the Netherlands Institute for Space Research (SRON). Since this simulator makes a few simplifying assumptions such as circular orbits and tiled ground pixels, satellite and instrument parameters were slightly modified to preserve essential parameters. In particular, orbit periods were calculated to match a
5   given cycle duration and length. The period then determines the altitude and inclination of a circular, sun-synchronous orbit. The altitude of the circular orbits is slightly larger than the typically used mean altitude for elliptic orbits. Since the altitude affects the size of the ground pixels and the width of the swath, field of view and along-track sampling time were set to match exactly the prescribed pixel size at sub-satellite point as well as the prescribed swath width. As a result, the number of
10   across-track pixels for Sentinel-5 did not match exactly the number of pixels for the real Sentinel-5 instrument.

Tables 2 and 3 summarize the orbits and viewing geometries of the two satellites. The CO2M satellite is assumed to have a 250-km wide swath and 11-day repeat cycle. Within the 11-day cycle, the instrument provides nearly global spatial coverage (Fig. 2). For locations in the SMARTCARB model domain, either one or two overpasses occur during the 11-day repeat cycle depending on the equator starting longitude of the satellite. Sentinel-5 has a much wider swath of 2650 km resulting in
15   near-daily global coverage (not shown).

$XCO_2$, CO and $NO_2$ column densities were sampled along the satellite swath for one year using the tracers from the COSMO-GHG simulations. For CO2M, $XCO_2$, CO and $NO_2$ columns were mapped onto the $2\,km \times 2\,km$ size pixels along the 250-km wide swath.

For Sentinel-5, CO and $NO_2$ columns were sampled with up to $7.5\,km \times 7.5\,km$ resolution along the 2670-km wide swath
20   of the Sentinel-5 instrument. Due to the wide swath, the pixel sizes grow towards the edge of the swath. In this study, only the spatial overlap between Sentinel-5 and CO2M were of interest, because Sentinel-5 was used for detecting the $CO_2$ emission plumes inside the swath of CO2M.

**8**

**Table 4.** Instrument uncertainty scenarios. VEG50 refers to a reference scene with a surface albedo of a vegetated surface and a solar zenith angle (SZA) of $50°$.

| Scenario name | Species | Satellite(s) | Reference noise ($\sigma_{VEG50}$ / $\sigma_{ref}$) absolute[1] | relative[1] |
|---|---|---|---|---|
| $CO_2$ low noise | $CO_2$ | CO2M | 0.5 ppm | - |
| $CO_2$ medium noise | $CO_2$ | CO2M | 0.7 ppm | - |
| $CO_2$ high noise | $CO_2$ | CO2M | 1.0 ppm | - |
| $NO_2$ low noise | $NO_2$ | CO2M | $1.0 \times 10^{15}$ molec. cm$^{-2}$ | 15% |
| $NO_2$ high noise | $NO_2$ | CO2M | $2.0 \times 10^{15}$ molec. cm$^{-2}$ | 20% |
| $NO_2$ Sentinel-5 | $NO_2$ | Sentinel-5 | $1.3 \times 10^{15}$ molec. cm$^{-2}$ | 20% |
| CO low noise | CO | Sentinel-5 and CO2M | $4.0 \times 10^{17}$ molec. cm$^{-2}$ | 10% |
| CO high noise | CO | Sentinel-5 and CO2M | $4.0 \times 10^{17}$ molec. cm$^{-2}$ | 20% |

[1] whichever is larger

### 2.1.3 Instrument error characteristics

The error characteristics of the $CO_2$, $NO_2$ and CO instruments were specified in collaboration with ESA based on previous studies for Carbonsat (Buchwitz et al., 2013) and on performance requirements for Sentinel-5 (Ingmann et al., 2012). For the instruments on the CO2M satellites, two or three scenarios were included in order to cover a realistic range between more or less demanding instruments. Table 4 summarizes the single sounding precision of the different instruments.

For XCO$_2$, three different uncertainty scenarios were considered which relate back to the performance estimates derived for the Carbonsat mission concept. Carbonsat was a $CO_2$ imaging spectrometer proposed for ESA's eighth Earth Explorer mission with specifications similar to those of the CO2M mission. A detailed error budget for Carbonsat was presented in the Carbonsat Report for Mission Selection (ESA, 2015). In the LOGOFLUX study, error parameterization formulas (EPF) for random and systematic errors were developed, which account for errors introduced by solar zenith angle (SZA), surface reflectance in the near infrared (NIR) and shortwave infrared (SWIR-1), cirrus clouds, and aerosol optical depth (Buchwitz et al., 2013).

Here, the same EPFs were adopted but only applied to compute random errors. These were calculated based on SZA and surface reflectance in the NIR and SWIR-1 band. Surface reflectances were taken from the MODIS MCD43A3 product (Version 006) at 1 km spatial resolution (Schaaf and Wang, 2015). A detailed consideration of cirrus clouds and aerosols and their impact on systematic errors was outside the scope of the study as it would have required the collection and processing of a large amount of additional data. The possible impact of not considering systematic errors will briefly be discussed in Sect. 4.

The random error calculated with the EPFs for the so-called vegetation-50 scenario (VEG50, i.e. vegetation albedo and SZA of $50°$) is about 1.5 ppm. In the model domain, mean random errors are slightly smaller at 1.3 ppm. To obtain random errors for the three instrument scenarios with $\sigma_{VEG50}$ of 0.5, 0.7 and 1.0 ppm, the computed errors were divided by 3.0, 2.14 and 1.5, respectively.

For $NO_2$ VCDs, the overall uncertainties are due to (a) measurement noise and spectral fitting affecting the slant column densities, (b) uncertainties related to the separation of the stratospheric and tropospheric column and (c) uncertainties in the

auxiliary parameters used for air mass factor (AMF) calculations such as clouds, surface reflectance, a priori profile shapes and aerosols (Boersma et al., 2004). The total uncertainties are dominated by uncertainties from spectral fitting for background pixels and by uncertainties in AMF calculations for polluted pixels, respectively. Typical spectral fitting uncertainty of previous instruments such as OMI were of the order of $1\text{-}2\times10^{15}$ molecules cm$^{-2}$ and AMF uncertainties of the order of 15-20%. These ranges were used to define two different scenarios for a possible CO2M $NO_2$ instrument (Table 4). For the Sentinel-5 UVNIS instrument, we assumed a relative uncertainty of 20% and a minimum uncertainty of $1.3\times10^{15}$ molecules per cm$^{-2}$. In the presence of clouds, the reference noise was increased using the empirical formula developed by Wenig et al. (2008). For a cloud fraction of 30%, random noise is approximately doubled.

For CO VCDs, the total uncertainty depends on the (a) fitting noise and (b) a priori CO and $CH_4$ profiles and (c) surface reflectance, aerosols and clouds. We assumed a single sounding precision of $4.0\times10^{17}$ molecules per cm$^2$ and a relative precision of 10% and 20% for both Sentinel-5 and the CO2M mission.

### 2.1.4 Cloud filtering

Satellite observations require filtering for clouds, which significantly reduces the number of observations available for plume detection. For the $CO_2$ product, we removed all $CO_2$ pixels with cloud fractions larger than 1%, because the $CO_2$ requires rigorous cloud filtering (Taylor et al., 2016). The $NO_2$ retrieval can tolerate larger errors and is therefore less sensitive to clouds. For the $NO_2$ product, we used a cloud threshold of 30% as often applied in satellite $NO_2$ studies (e.g. Boersma et al., 2011). For CO, a cloud threshold of 5% was used, which is motivated by the cloud threshold used for the MOPITT CO product (Deeter et al., 2017; MOPITT Algorithm Development Team, 2017). All cloud information in this study was obtained from Previous studies used the MODIS cloud mask product available at 1 km resolution (Ackerman et al., 2017) for masking cloudy $CO_2$ observations (Buchwitz et al., 2013; Pillai et al., 2016). Since CO and $NO_2$ observations can tolerate larger cloud fractions, a cloud fraction product would be needed for masking pixels with different thresholds but the MODIS cloud product is only available at 5 km resolution (Platnick et al., 2017). Therefore, we used total cloud fractions computed by COSMO-GHG, i.e. the same model as used for the tracer transport simulations., that are available at model resolution. COSMO-GHG computes total cloud fraction from layer cloud fractions assuming minimum overlap. The differences between clouds masks derived from COSMO-GHG and MODIS products and their effects on data yield are discussed in Section 4.1.

## 2.2 Plume detection algorithm

### 2.2.1 Algorithm

We developed a new but simplified plume detection algorithm that uses a statistical test to detect signal enhancements which are significant with respect to instrument noise and variability in background levels. The plume is then identified as a coherent structure of significant pixels. The algorithm involves three processing steps as laid out in Fig. 3.

The first step of the the plume detection algorithm finds satellite pixels with $CO_2$, CO or $NO_2$ values significantly larger than the background field using a statistical Z-test, for which the distribution of the test statistics can be approximated by a
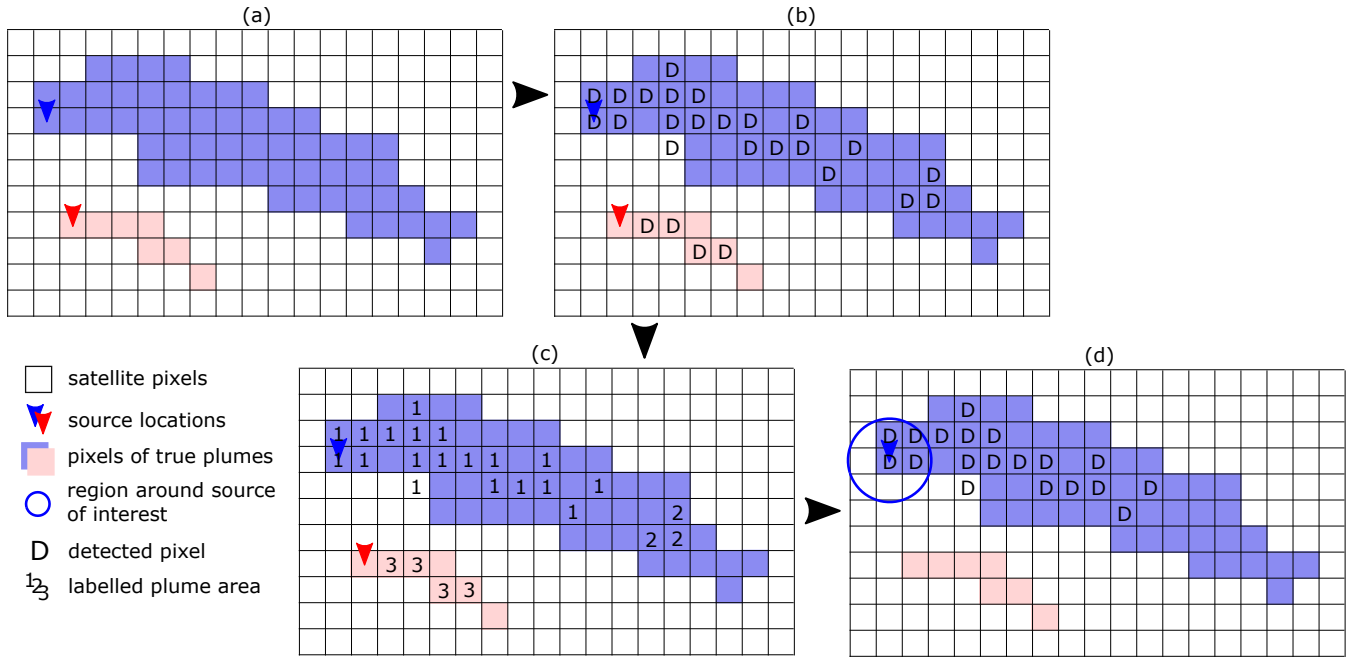
**Figure 3.** Schematic of the processing steps of the plume detection algorithm: (a) A large (~~red~~blue) and small plume (~~blue~~red) with sources marked by arrows are located within the satellite overpass. (b) Pixels detected based on Z-test are marked with "~~T~~D" ~~for True~~. (c) Connected pixels are given a unique number each denoting a different plume. (d) All plumes not connected with the source of interest (blue circle) are rejected.
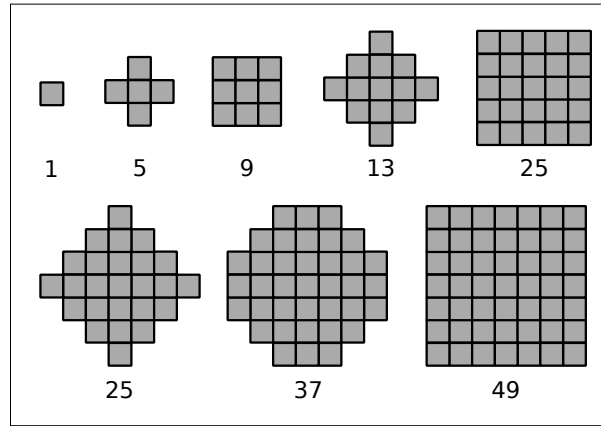


**Figure 4.** Examples of neighborhoods with different sizes ($n_s$) used for calculating the local mean.

normal distribution (e.g. v. Storch and Zwiers, 2003). The Z-test computes a z-value given by

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

**11**

where $x$ is an observation from a population with mean value $\mu$ and standard deviation $\sigma$. The value $x$ is considered significantly larger than $\mu$ when the z-value is greater or equal a critical value. The critical value is calculated from the inverse cumulative distribution function of the normal distribution for a probability $q$. The probability that $x$ is not significantly larger than $\mu$ is the p-value ($p = 1 - q$).

A key feature of the algorithm is that the ~~signal~~ trace gas observation of a single pixel is replaced by a spatial average of pixels in a defined neighborhood. This allows identifying weak plumes with signals of individual pixels well below instrument noise, but also bares the risk of diluting the signal at the plume edges. Figure 4 shows examples of neighborhoods with different sizes $n_s$ that have been used for testing the algorithm. A large neighborhood results in a stronger smoothing and may therefore produce a larger number of false positives, i.e. pixels outside of the plume that are wrongly assigned to the plume. An ideal neighborhood size balances the need for sensitive plume detection with the requirement of a low fraction of false positives.

Whether a signal enhancement is detectable is primarily determined by the z-value (Eq. 1) which can also be interpreted as the signal-to-noise ratio (SNR) at a spatially smoothed satellite pixel. Thereby, the signal is the enhancement above background due to the plume, and the noise is composed of both instrument noise and spatial variability in the background. Since the background, i.e. the trace gas field in the absence of the plume, can not directly be observed, it needs to be estimated either from the observable trace gas field surrounding the plume or from a climatology or a model. Instrument noise and background variability can have both random and systematic components. The random component would reduce with the inverse square root of the number of valid pixels $n$ ~~of~~ in the neighborhood, whereas the systematic error would be approximately independent of $n$.

~~The SNR is thus computed as~~ The number of valid pixels $n$ can be smaller than the size of the neighborhood $n_s$ when pixels are missing, e.g. due to clouds or at the boundary of the satellite swath. Therefore, the z-value or the SNR can be calculated as follows:

$$\mathrm{SNR} = \frac{X_\mathrm{obs} - X_\mathrm{bg}}{\sqrt{\frac{\sigma_\mathrm{rand}^2}{n} + \sigma_\mathrm{sys}^2}} \tag{2}$$

where $X_\mathrm{obs}$ are the ~~observed trace gas values~~ spatially averaged satellite observations, $X_\mathrm{bg}$ is the estimated background value, and $\sigma_\mathrm{rand}$ and $\sigma_\mathrm{sys}$ are the random and systematic errors, respectively.

~~The first step of the the plume detection algorithm finds satellite pixels with , CO or values significantly larger than the background field by using a statistical Z-test, for which the distribution of the test statistics can be approximated by a normal distribution (e.g. v. Storch and Zwiers, 2003). The Z-test calculates the SNR~~ Equation (2) can be calculated for each satellite pixel and ~~compares it to the z-value $z(q)$ for~~ compared with the critical value to determine which $X_\mathrm{obs}$ ~~is~~ are significanlty larger than $X_\mathrm{bg}$ ~~with probability $q$:~~

$$\mathrm{SNR} \geq z(q).$$

$\sim$

To compute the SNRs from the satellite image, we computed observed values $X_{\text{obs}}$ as the local means of a neighborhood of size $n_s$ (Fig. 4). For sake of simplicity, the background $X_{\text{bg}}$ in this study was estimated from the X_BER_BG and X_PP_BG model tracers for a $200{\times}200$ km$^2$ square centered on the city of Berlin and a $100{\times}100$km$^2$ square centered on each power plants. The implications of this assumption are discussed in Sect. 4.

5      For random and systematic errors we assumed that the instrument uncertainty is purely random and that the background variability is purely systematic. For the instrument uncertainty, random errors as listed in Table 4 were used, which reduce with the number of valid pixels in the local neighborhood $n$ due to the inverse scaling by the number of pixels. The background uncertainty $\sigma_{bg}^2$ was computed from the spatial variance of the background model tracer in a fixed domain surrounding the source as described above. Since it is assumed to be systematic, it does not decrease with the size of the neighborhood.

10      The result of the Z-test is a binary image with "true" values where pixels are significantly enhanced above the background and "false" values where they aren't (Fig. 3b). Since the local mean can still be computed for missing center pixels - using neighboring pixels - missing pixels can also be detected as enhanced above the background.

     In the second step, pixels that are enhanced ("true") and connected are assumed to belong to the same plume. ~~Regions We label regions~~ of connected pixels using a standard labeling algorithm. Neighboring pixels are identified using a Moore

15 neighborhood, where each pixel has eight potential neighbors~~, and are labelled with~~. Each region is assigned a unique integer value (Fig. 3c).

     Finally, in the third step, all connected regions that do not intersect with the source region are removed leaving only regions that overlap with the source of the plume (Fig. 3d). For cities, the "source region" is defined by a circle with a radius of 15 km and for point sources by a circle with a radius of 5 km. The last step may remove regions that are part of the real plume but

20 separated from the source by weak signals or missing values (e.g. the region labeled "2" in Fig. 3c).

### 2.2.2    Performance evaluation

The plume detection algorithm was applied to the $CO_2$ plumes of Berlin and Jänschwalde. ~~Its performance was evaluated against the~~ The detectability of $CO_2$ plumes was evaluated for the different instrument scenarios by comparing the detected plume with the "true plume" defined by the field of the $CO_2$ tracer released by the respective source (CO2_BER in case of

25 Berlin, CO2_PP for Jänschwalde) above a low threshold. For Jänschwalde, the ~~true plume~~ performance had to be additionally ~~identified~~ evaluated by visual inspection due to frequent overlaps with the plumes of other power plants also contained in the CO2_PP tracer.

     To evaluate the performance of the Sentinel-5 with respect to detecting plumes as observed by CO2M, detected pixels had to be projected onto the pixels of the CO2M instrument. Sentinel-5 pixels were thus only used over the swath of the CO2M

30 satellite rather than over the whole swath of Sentinel-5.

     When using $NO_2$ or CO for plume detection, the performance was assessed by comparing the detected pixels with the true $CO_2$ plume rather than the true plume of the auxiliary gas. In this way, the degree of congruence between the $CO_2$ and auxiliary trace gas plumes was considered as well.

For the evaluation, we computed true positives (TP), false positives (FP) and the positive predictive value (PPV = TP/(TP+ FP)) (Ting, 2010). A good algorithm should have a much smaller number of FP than TP and, therefore, a PPV close to 1. To remove the impact of different cloud thresholds, TP, FP and PPV were only computed for cloud-free pixels using a cloud threshold of 1%, because we are primarily interested in valid $CO_2$ observations that can be used for estimating $CO_2$ emissions.

5

## 3 Results

### 3.1 Coverage and potential for plume detection

In this section, the potential for plume detection is analyzed based on the simulated tracers emitted by the source of interest. These "true plumes" will be used in the following sections as reference to evaluate the performance of the plume detection
10 algorithm. They can be interpreted as the maximum number of plumes detectable by a perfect, noise-free instrument.

The frequency with which the $CO_2$ plumes of a given source can be observed depends on how often a satellite passes over the source, how often the $CO_2$ signal is larger than the threshold and how often cloud-free conditions dominate during the overpass. We define an overpass as an intersection between the satellite swath and the "source region" as specified above. The number of overpasses scales with the number of satellites and the swath width of the instrument. The scaling with the number
15 of satellites is not trivial, however, since individual satellites may pass over the source either once or twice during the 11-day repeat cycle depending on the satellite's equator starting longitude (see Fig. 2).

The total number of overpasses per satellite is either 34 or 66 per year, depending on whether the satellite has one or two overpasses per 11-days repeat cycle. Since one out of four satellites has only one overpass per 11 days, the number of overpasses per year roughly scales with a factor 1.75 times the number of satellites times the number of 11-day periods per year (about
20 33). A constellation of six satellites covers the model domain nearly daily.

To define the extent of a plume in the satellite image, we have to set a signal threshold for the tracer field (XCO2_BER for Berlin) above which a pixel is considered as belonging to the plume. A possible threshold is the value at which the signal would become larger than the variability of the background, i.e. where the signal is larger than the standard deviation of the background. Based on the time series of standard deviations of the model background tracer (XCO2_BER_BG for $CO_2$)
25 computed for a $200{\times}200$ km$^2$ square centered on the city of Berlin (Fig. 7c, f and i), we defined a threshold of 0.05 ppm for $XCO_2$, $0.2{\times}10^{15}$ molec. cm$^{-2}$ for $NO_2$ and $0.06{\times}10^{17}$ molec. cm$^{-2}$ for CO approximately corresponding to the minimum of the standard deviations. Note that these thresholds are significantly smaller than the noise level of the instruments.

A plume was defined as the collection of pixels for which the signal is larger than the threshold. However, we also required that Berlin is inside the swath of the instrument to be able to unambiguously assign a plume to the city. Furthermore, we
30 removed parts of plumes that re-entered the swath after leaving it, because it is often not possible to correctly assign these parts to their source.

Since a satellite image can be obscured by clouds, we need to define how many pixels are needed to make up a "useful" plume. This number depends on the application. For example, to estimate emissions of cities, we require that the plume must
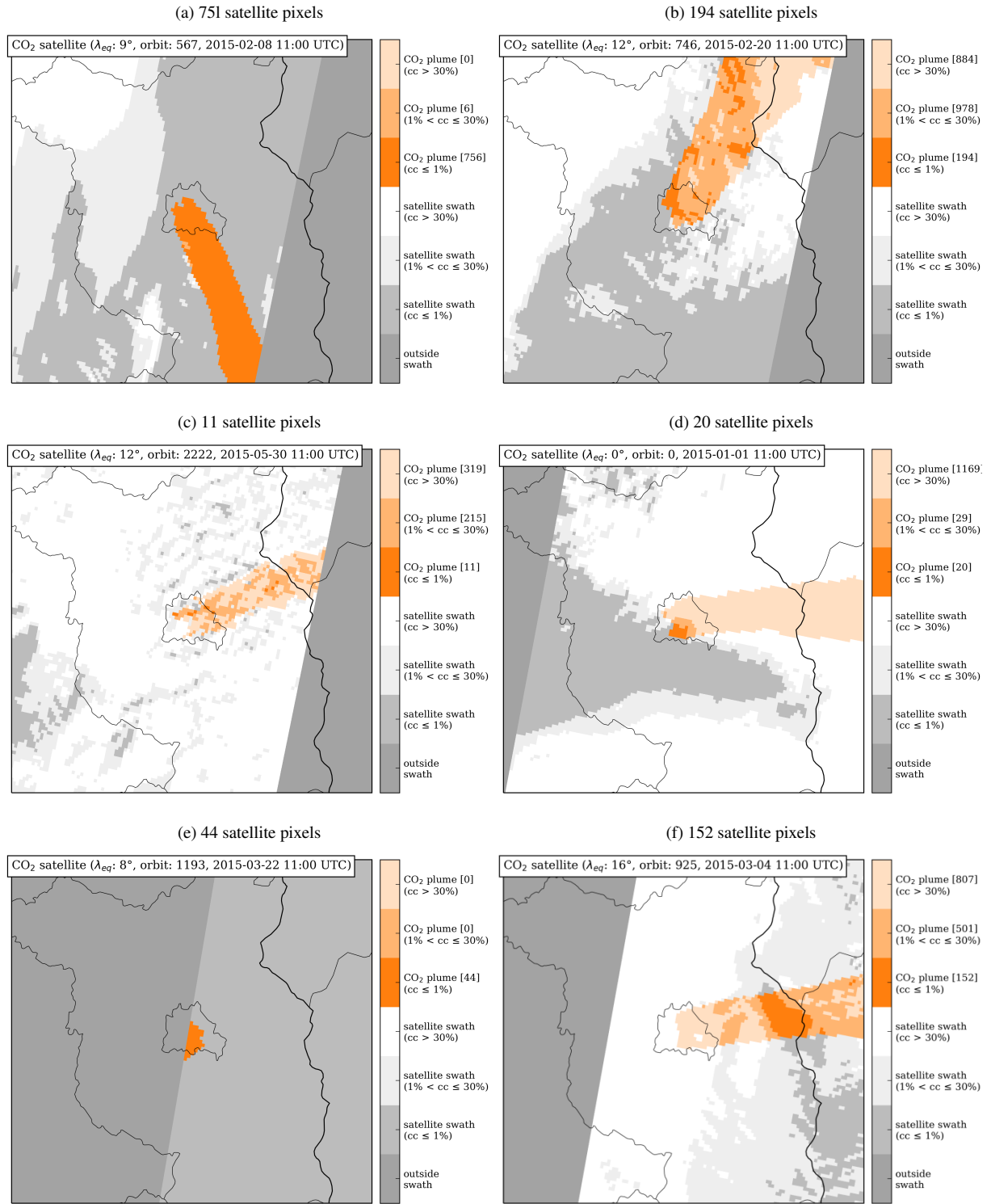
**14**

**Figure 5.** Examples of [true](#) XCO$_2$ plumes of Berlin (XCO$_2$ signal > 0.05 ppm) with different cloud cover fractions (cc). The numbers of XCO$_2$ pixels are shown for cloud fractions ≤1%. (a-d) Plumes with increasing cloud fraction, (e) plume close to the edge of the swath, (f) plume without cloud-free CO$_2$ observations connected to Berlin.

**15**

extend beyond the city limits to contain emissions from the whole city area. The cross-wind diameter of Berlin's $CO_2$ plume is typically about 20 km or 10 satellite pixels, which is roughly the diameter of the part of the city with the highest emissions. To cover at least the whole city area, we only consider $CO_2$ plumes with at least 100 cloud-free $CO_2$ pixels to be useful. For the power plants, the cross-wind diameter is less than five pixels near the source. Therefore, 10 pixels were used to define the minimum number for a useful plume in this case. It should be noted that this number of pixels is not necessarily sufficient for estimating the emissions of a source with certain accuracy, which depends, among others, on instrument precision, meteorology and source strength. Nonetheless, detecting the full crosswind diameter is the minimum requirement, for example, for flux-based inversion methods (e.g. Krings et al., 2013; Reuter et al., 2019). The number of detected pixels is a useful measure for comparing the detectability of $CO_2$ plumes with $CO_2$, CO and $NO_2$ observations, because source strength and meteorology are the same for a given source.

For a signal threshold of 0.05 ppm, Berlin's $CO_2$ plume has always more than 100 pixels above the signal threshold. However, many plumes are partly or fully covered by clouds significantly reducing the number of useful plumes. Figure 5a-d presents examples of $CO_2$ plumes under different cloud conditions with an increasing fraction of cloudy pixels. Figures 5c and d show examples with plumes of only 11 and 20 pixels, much smaller than the area of the city. On the other hand, the 100-pixel threshold does not necessarily remove swaths with plumes in broken clouds (e.g. Fig. 5b), for which it will also be challenging to estimate emissions, because adjacent cloudy pixels increase the $XCO_2$ uncertainty (Taylor et al., 2016). The number of plumes with at least 100 pixels is also reduced when the source is close to the edge of the swath and winds are pushing the plume out of the view of the satellite (e.g. Fig. 5e). These overpasses occur every 11 days due to the repeat cycle of the satellite. As a result, orbits with plumes near the edge of the swath can have up to 20% less useful plumes.

Figure 6 presents the number of useful city plumes (>100 pixels) per month for $CO_2$, $NO_2$ and CO for constellations of one to six satellites. Plumes without cloud-free observations over the source region (e.g. Fig. 5f) were removed, because they cannot be detected by the algorithm used in this study. A constellation of six satellites observes only 50±5 $CO_2$ plumes within one year despite almost daily overpasses due to the small number of days with low cloud fractions. Except for February, which was an unusually sunny month in 2015, there is a clear tendency of higher cloud fractions and correspondingly fewer plume observations in winter than in summer. The standard deviations shown in the figures as vertical black bars were estimated from the scatter of observable plumes using satellites with different equator starting longitudes. The presence of clouds thus reduces the opportunity for plume detection by a factor as a large as 5 to 6 over the city of Berlin. The number of observable $NO_2$ plumes per year (108±8) is about twice as large as for $CO_2$, which is primarily due to the larger cloud threshold of 30%. For CO the number of observable plumes per year is 58±5. The average number of plumes per satellite and year is thus about 8 (range: 3-13), 9 (4-15) and 17 (7-23) for $CO_2$, CO and $NO_2$, respectively.

The number of observable plumes varies strongly between the individual satellites of a constellation, because the number of cloud-free days per year is quite small and the overpass days are different for different equator starting longitudes. Since satellites are equally spaced in orbit, changing the number of satellites changes the starting longitudes and overpass days of the satellites. As a consequence, the number of observable plumes per constellation can also fluctuate strongly. According to Fig. 6, for example, a constellation of two satellites seems almost equivalent to a constellation of three, but this result is merely

**Figure 6.** Number of cloud-free plumes of the city of Berlin with at least 100 pixels per month for (a) $CO_2$, (b) $NO_2$ and (c) CO. The cloud threshold is 1% for $CO_2$, 30% for $NO_2$ and 5% for CO observations. Error bars are obtained by comparing all available satellites. The number of expected plumes per satellite is 8, 17 and 9 for the $CO_2$, $NO_2$ and CO instrument, respectively.

a consequence of the fact that cloud cover was often large during these overpasses and Berlin was at the edge of the swath for the satellite with a starting longitude of 8°. The result would be different for another starting longitude of the first satellite, another city, or another year.
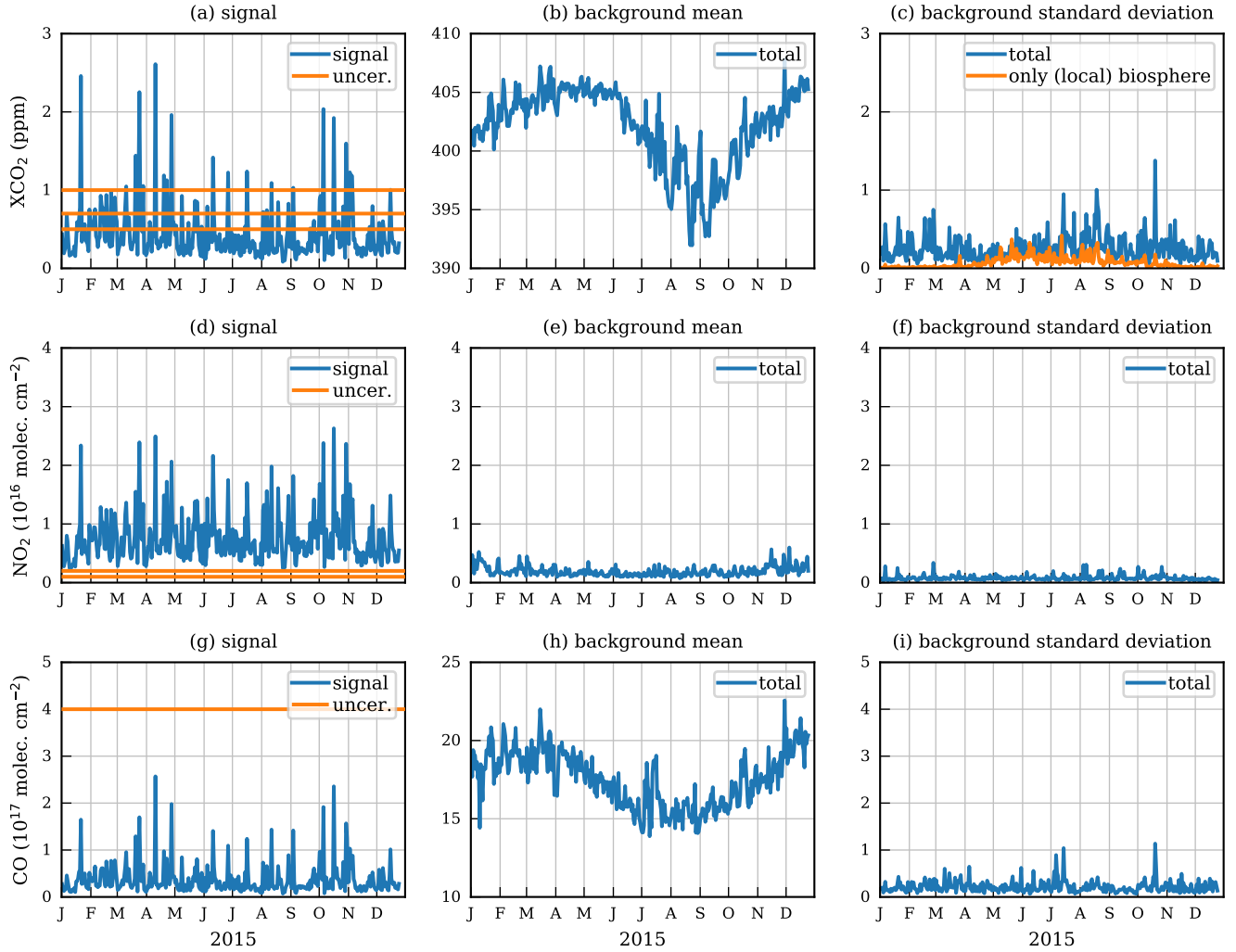
**Figure 7.** Time series of $CO_2$, $NO_2$, CO plume signals (left column), mean backgrounds (middle) and standard deviations of backgrounds (right) for Berlin. Signals are the largest local mean values of the X_BER model tracer using a 37-pixel-neighborhood. Background means and standard deviations were obtained from the X_BER_BG tracer for a $200 \times 200$ km$^2$ square centered over Berlin. Reference uncertainties ($\sigma_{\mathrm{VEG50/ref}}$) corresponding to the different instrument scenarios are shown as horizontal lines for comparison.

## 3.2 Signal-to-noise ratios

The key measure that determines the detectability of a $CO_2$ plume is the SNR (Eq. 2), which compares the amplitude of the plume signal to the instrument noise and the variability of the background. SNRs provide a first indication of an instrument's suitability for detecting a plume.

Time series of the $CO_2$, $NO_2$ and CO plume signals were computed from the X_BER and X_PP tracers for Berlin and the power stations Jänschwalde and Lippendorf at the overpass time of CO2M, i.e. about 11 UTC. The signals were computed as

**Table 5.** 5th, 50th and 95th percentile of $CO_2$, $NO_2$ and CO signals of Berlin as well as Jänschwalde and Lippendorf power stations.

| Species | 5th | 50th | 95th |
|---|---|---|---|
| Berlin: | | | |
| $CO_2$ (ppm) | 0.16 | 0.33 | 1.03 |
| $NO_2$ ($10^{16}$ molec. cm$^{-2}$) | 0.30 | 0.63 | 1.56 |
| CO ($10^{17}$ molec. cm$^{-2}$) | 0.12 | 0.27 | 0.96 |
| Jänschwalde power station: | | | |
| $CO_2$ (ppm) | 1.28 | 2.69 | 6.64 |
| $NO_2$ ($10^{16}$ molec. cm$^{-2}$) | 2.09 | 4.30 | 10.0 |
| CO ($10^{17}$ molec. cm$^{-2}$) | 0.55 | 1.14 | 2.88 |
| Lippendorf power station: | | | |
| $CO_2$ (ppm) | 0.53 | 1.29 | 3.73 |
| $NO_2$ ($10^{16}$ molec. cm$^{-2}$) | 0.85 | 2.03 | 5.37 |
| CO ($10^{17}$ molec. cm$^{-2}$) | 0.03 | 0.08 | 0.22 |

**Table 6.** Median signal-to-noise ratios for signals of Berlin, Jänschwalde and Lippendorf using different uncertainty scenarios. The signals were computed as largest local mean values using a local neighborhood size $n_s$ of 37 and 5 for cities and power stations, respectively.

| Scenario name | Signal-to-noise ratio | | |
|---|---|---|---|
| | Berlin | Jänschwalde | Lippendorf |
| $CO_2$ low noise | 1.4 | 10.4 | 4.3 |
| $CO_2$ medium noise | 1.4 | 8.0 | 3.5 |
| $CO_2$ high noise | 1.2 | 5.8 | 2.6 |
| $NO_2$ low noise | 9.0 | 14.3 | 8.8 |
| $NO_2$ high noise | 8.4 | 10.8 | 7.5 |
| CO low/high noise | 0.4 | 0.6 | 0.0 |

maximum values of the local means within the source region, i.e. a circle with 15 or 5 km radius. Thereby, the local means were computed with a neighborhood $n_s$ of size 37 for Berlin and 5 for the power stations (Fig. 4). ~~These sizes~~ A large neighborhood reduces the random noise of the measurements and therefore allows detecting smaller signals. On the other hand, a large neighborhood will include background values in the computation of the averages at the plume edges and reduce the signal.

5 The sizes used here roughly correspond to the typical diameters of the $CO_2$ plumes from Berlin (about 15 km) and the power stations (about 6 km), respectively, and were also found most suitable for the plume detection algorithm, because they maximize TPs without reducing PPVs too much (see also Kuhlmann et al., 2019). The results for Berlin are presented in Fig. 7 for the three trace gases. The figure compares the daily plume signals (left panels) to the daily mean background values (middle) and their spatial variability (right). The 5th, 50th and 95th percentiles of the time series are summarized in Table 5 for Berlin as

10 well as for the power stations. The signals have a large range due to the variability of emissions (e.g. lower during weekends) and meteorology. The $CO_2$ and $NO_2$ signals of the power stations are between five and ten times larger than those of Berlin.

The CO signal of Lippendorf, on the other hand, is smaller than the signal of Berlin. The power plants produce strong local enhancements easily detectable by the $CO_2$ satellite, but the corresponding plumes are much narrower than those of Berlin.

Figures 7b and c present the spatial means and standard deviations of the background around Berlin. Background $XCO_2$ has a strong annual cycle with an amplitude of about 16 ppm. Since the $XCO_2$ plume signal of Berlin is typically only about 0.2 to 1.0 ppm, it is critical to accurately estimate the background $XCO_2$ value in Eq. (??2). The spatial variability $\sigma_{bg}$ of the background, on the other hand, is typically only of the order of a few tenths of a ppm. Despite higher $XCO_2$ in winter than in summer, the variability is somewhat larger in summer due to stronger biospheric activity in combination with lower average wind speeds, especially in July and August. Large peaks in the background variability are often caused by plumes from other anthropogenic sources such as the power stations in the south-east of Berlin (Fig. 1).

For $NO_2$, the annual cycle of the background is relatively constant for our idealized $NO_2$ tracer with a constant lifetime of 4 hours (Fig. 7e and f). In reality, the lifetime will likely be longer and the variability correspondingly higher in winter. The $NO_2$ signal of Berlin is significantly larger than the background and its variability. Similar to $CO_2$, the CO time series has a strong annual cycle with an amplitude of about $5 \times 10^{17}$ molec. $cm^{-2}$ (Fig. 7h and i) requiring again an accurate estimation of the background. The standard deviation of the background is about half of the CO signal.

Table 6 summarizes the median of all SNRs of Berlin and the two power stations for the different satellite instrument scenarios that have been computed from the time series of highest signals. To understand the numbers, it should be noted that a plume pixel would be detectable when the SNR is larger than 2.3, i.e. $z(q) = 2.3$ for $q = 99\%$. For Berlin, the $CO_2$ SNRs are below this detection limit for all noise levels while $NO_2$ SNRs are above the limit. For the two power stations, SNRs are above the detection limit both for the $CO_2$ and $NO_2$ instrument scenarios, but SNRs for the $NO_2$ instrument scenarios are always larger.

Based on the SNRs, the $NO_2$ plumes should be well detectable. For Berlin, the detection of the $CO_2$ plume with the $CO_2$ instrument will often be challenging due to low SNRs. The CO SNRs are always much smaller than those for $CO_2$, making a CO instrument with the given specifications little suitable for the purpose of plume detection. In the following, we therefore only investigate the potential benefit of auxiliary $NO_2$ observations.

## 3.3 Plume detection algorithm

The plume detection algorithm was applied to the $CO_2$ plumes of Berlin and Jänschwalde for different instrument scenarios. The probability $q$ was set to 99% and neighborhood sizes of 37 and 5 were selected for Berlin and the power stations, respectively. In the case of Sentinel-5, the corresponding neighborhood sizes were set to 5 and 1 due the larger pixels of this satellite. Based on an analysis of the positive predictive values (PPV), these neighborhood sizes were found most suitable for detecting the city and power plant plumes (Kuhlmann et al., 2019). For Berlin, twenty synthetic satellite images were created for each single overpass with different patterns of random noise. The plume detection algorithm was subsequently applied to each image and the results were averaged to obtain more robust results independent of the selected noise pattern. For Jänschwalde, only one synthetic satellite image was created for each overpass, because no model tracer was available to compare the results with a true plume.
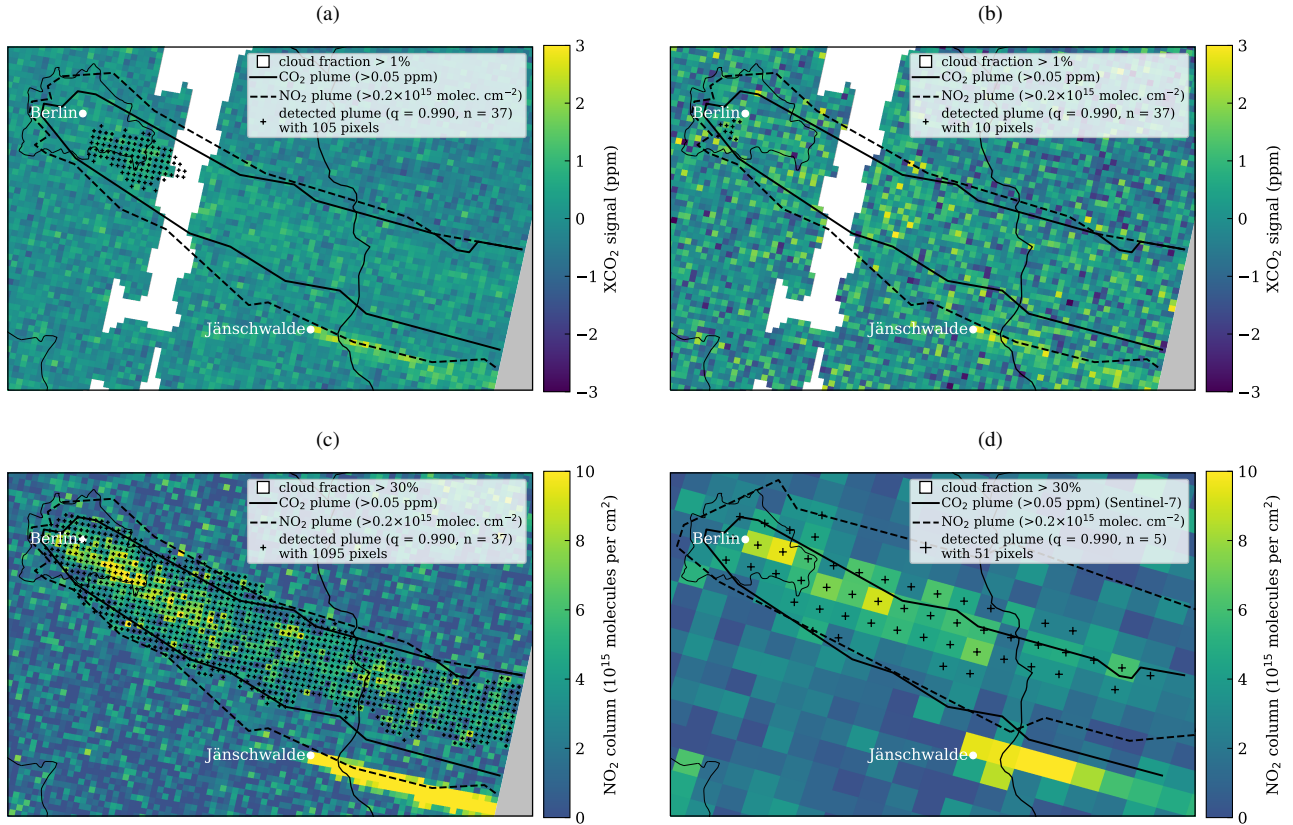
**Figure 8.** Example of plume detection with CO2M's $CO_2$ and $NO_2$ instrument and Sentinel-5's $NO_2$ instrument on 21 April 2015. Significant pixels detected by the algorithm are highlighted as black dots. The outlines of the true $CO_2$ and $NO_2$ plumes based on the $X_{BER}$ tracers are overlaid as solid and dashed lines, respectively. (a) Low-noise $CO_2$ instrument. (b) high-noise $CO_2$ instrument. (c) High-noise $NO_2$ instrument on the CO2M satellite. (d) $NO_2$ instrument on Sentinel-5.

### 3.3.1 Examples of detected plumes from Berlin

Figure 8 shows the $CO_2$ and $NO_2$ plumes of Berlin on 21 April 2015 observed by CO2M and Sentinel-5 for different instrument scenarios. The outlines of the real plumes are overlaid as solid and dashed lines for $CO_2$ and $NO_2$, respectively. Since the $CO_2$ instruments have a lower cloud threshold, a band of cirrus clouds is obscuring the plume in the $CO_2$ observations but not

5     in $NO_2$. Successfully detected pixels are shown as black crosses and the number of detected pixels (median of all 20 noise realizations) are presented in the legend. On average, a $CO_2$ instrument detects 116±46, 48±40 and 24±30 pixels with noise scenarios $\sigma_{\mathrm{VEG50}}$ of 0.5, 0.7 and 1.0 ppm, respectively (Fig. 8a and b). The number of true positives is slightly smaller having on average two false positive pixels. Consequently, the PPV is high, ranging between 0.85 and 0.99 for high and low noise, respectively.

10     For the $NO_2$ measurement the band of thin cirrus clouds is not an issue. The $NO_2$ instrument can therefore detect a much larger number of pixels, i.e., 1242±99 and 1203±155 in the case of the low and the high noise scenarios, respectively (Fig. 8c).
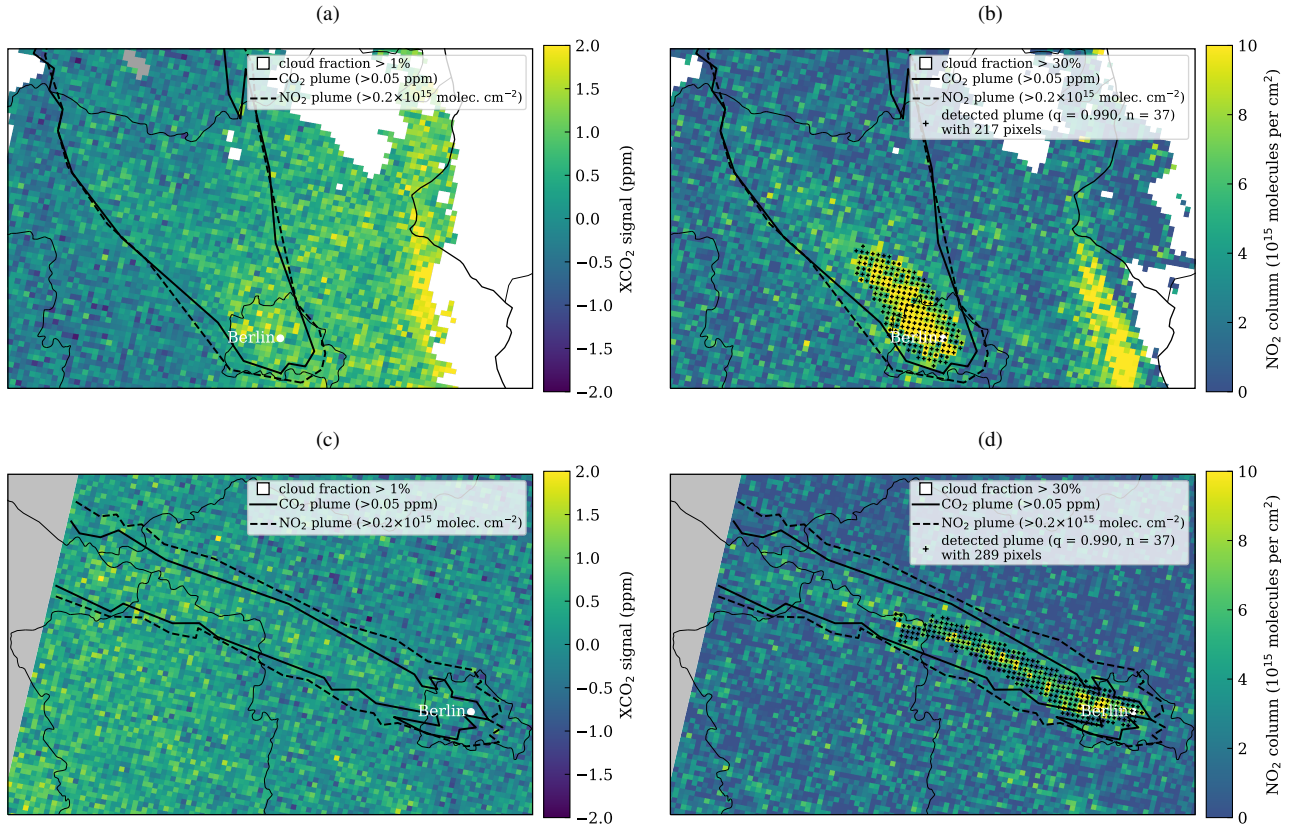
**Figure 9.** (a) Example of 27 February 2015 where plume detection with a $CO_2$ instrument fails because of pronounced horizontal gradients in the $CO_2$ background field. (c) Example of 2 July 2015 where the plume detection fails due to small $CO_2$ signals as a result of high wind speeds. In both cases, the plume can readily be detected with a high noise $NO_2$ instrument (panels b and d).

On average, the fraction of FP is relatively large and the PPV is only $0.80\pm0.05$ and $0.77\pm0.10$ for the low and high noise scenarios, respectively. The small PPV is caused by interference with the plume of Jänschwalde, which is just south of the plume of Berlin. For cases where no neighboring plumes have been detected falsely with the $NO_2$ instrument, the spatial match between $CO_2$ and $NO_2$ plumes is generally high, suggesting a high degree of spatial overlap between the $CO_2$ and $NO_2$

5 plumes.

The Sentinel-5 $NO_2$ instrument is also able to detect the $CO_2$ plume with $879\pm114$ CO2M pixels, but since it is measured two hours earlier, the $NO_2$ plume seen by Sentinel-5 (dashed line in Fig. 8d) is slightly shifted with respect to the $CO_2$ plume (solid line). As a consequence, the PPVs is low ($0.60\pm0.04$).

Figure 9 presents two examples where the $CO_2$ instrument fails to detect the $CO_2$ plume. In Fig. 9a, the $CO_2$ field has a

10 pronounced spatial gradient resulting in a high variance of the background. This gradient is not present in the much shorter lived trace gas $NO_2$ making it possible to detect the plume using an $NO_2$ instrument (Fig. 9b). Similar situations occur in roughly 20% of cloud-free swaths. Figures 9c and d show a second example where the $CO_2$ instrument cannot detect the
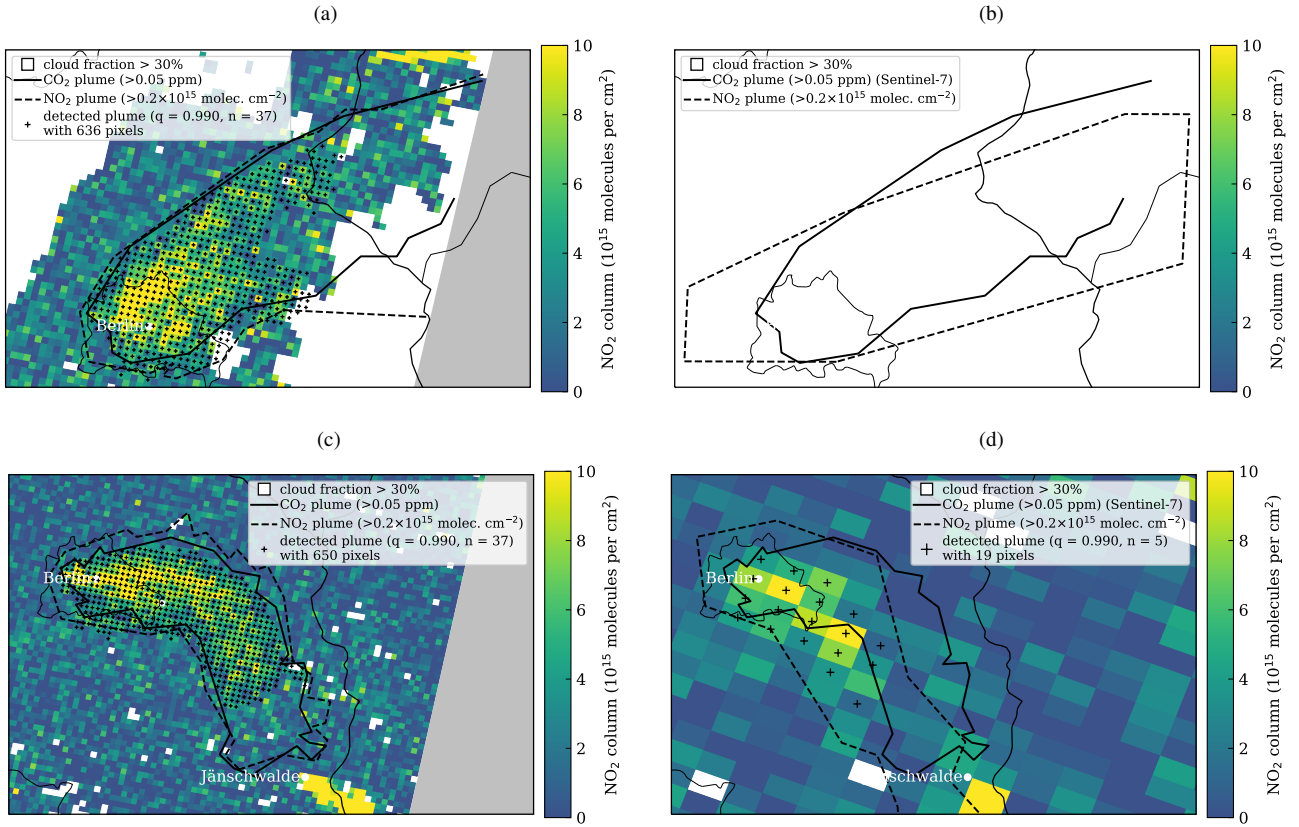
**Figure 10.** Examples of comparing plume detection between the Sentinel-5 and CO2M's $NO_2$ instruments (a, b) Due to time-lag of two hours, the scene is cloudy during the Sentinel-5 overpass but not during the CO2M overpass (3 December 2015). (c, d) The plume position clearly changed between the two overpasses (17 June 2015).

plume, because the signal is very weak due to strong winds. Owing to its better SNR, the $NO_2$ instrument is able to detect the plume also in this situation.

Figure 10 presents two examples comparing the $NO_2$ plume observed by Sentinel-5 to the $CO_2$ plume observed two hours later by CO2M satellite. In the first example (panels a and b), Sentinel-5 fails to detect any plume due to clouds, which have largely disappeared by the time of the CO2M overpass. In the second example, both Sentinel-5 and the CO2M satellite detect a plume of similar size, but the Sentinel-5 plume is significantly displaced due to changes in the prevailing winds between the two overpasses.

### 3.3.2 Number and sizes of detected Berlin plumes

To count the number of plumes detectable under the different instrument scenarios, we analyze the fifty plumes observed by a constellation of six satellites, which we had classified in Sect. 3.1 as being potentially "useful" based on the idealized tracer $X_{BER}$ having more than 100 pixels above a threshold of 0.05 ppm. Table 7 summarizes the results in terms of number and

**Table 7.** Number of $CO_2$ plumes from Berlin and 5th, 50th and 95th percentiles of the number of detected $CO_2$ pixels (TP if PPV $\geq 0.8$ and cloud fraction < 1%) for a constellation of six CO2M satellites or one Sentinel-5 satellite. The maximum number of detectable plumes would be 50, corresponding to all potentially useful plumes with at least 100 $CO_2$ pixels with values of the tracer $X_{BER}$ above a threshold of 0.05 ppm.

| Instrument scenario | Plumes with $\geq 100$ $CO_2$ pixels | | Plume size at percentile | | |
| --- | --- | --- | --- | --- | --- |
| | number | percentage (%) | 5th | 50th | 95th |
| $CO_2$ low noise | 16±1 | 32±3 | 0 | 5 | 323 |
| $CO_2$ medium noise | 10±1 | 20±3 | 0 | 3 | 261 |
| $CO_2$ high noise | 6±1 | 12±2 | 0 | 7 | 181 |
| $NO_2$ low noise | 34±1 | 68±2 | 52 | 294 | 600 |
| $NO_2$ high noise | 35±2 | 70±3 | 50 | 279 | 527 |
| $NO_2$ Sentinel-5 | 10±2 | 20±3 | 0 | 140 | 396 |

size of the detected plumes. A plume was only counted as detected when at least 100 $CO_2$ pixels were correctly detected (true positives) and when at least 80% of the detected pixels were true positives (PPV $\geq 0.80$). The PPV threshold was found useful for removing plumes interfering with others or plumes shifted due to the earlier overpass time of Sentinel-5.

Table 7 shows that the $CO_2$ instruments detect significantly fewer plumes than the $NO_2$ instruments. Depending on instru-
5  ment noise scenario, the $CO_2$ instruments detect plumes with more than 100 pixels with a success rate of only 12% to 32%, while for the $NO_2$ instruments the success rates are 68% to 70%. Surprisingly, the $NO_2$ instrument with low noise performs slightly worse than the high noise instrument. This is an artifact of the algorithm often detecting small plumes not related to emissions from Berlin in the case of a low-noise instrument. The Sentinel-5 $NO_2$ instrument detects 20% of the plumes, thus only half the success rate of the $NO_2$ instrument on the CO2M satellite. The main reason for this low success rate is the spatial
10  mismatch of the plumes due to the two hours difference in overpass times.

Figure 11 shows the number of plumes per month with at least 100 detected pixels for different constellations between one and six satellites for the $CO_2$ low noise and the $NO_2$ high noise scenario. The number of plume detections per month is small and therefore highly sensitive to the specific orbit configuration. For example, two satellites seem to detect more plumes than three, but this result is caused by an unfavorable orbit for observing Berlin for the constellation with three satellites and
15  unfavorable cloud cover as already discussed earlier. The standard deviation was estimated from the number of detectable plumes using satellites with different starting longitudes (i.e. east-west displacements of all orbits). The figure shows that the number of observed plumes generally increases with the number of satellites as expected, but statistical noise can mask the increase from one constellation to the next. The figure confirms the much lower success rates of the $CO_2$ instruments as compared to the $NO_2$ instruments as expected from the computed signal-to-noise ratios.

20  ### 3.3.3 Detection of plumes from power stations

There are six major power plants in the model domain: Jänschwalde, Boxberg, Schwarze Pumpe, Lippendorf, Turow and Patnow. Because no model tracer was defined for individual power plants but only for the sum of all of them, the true plume
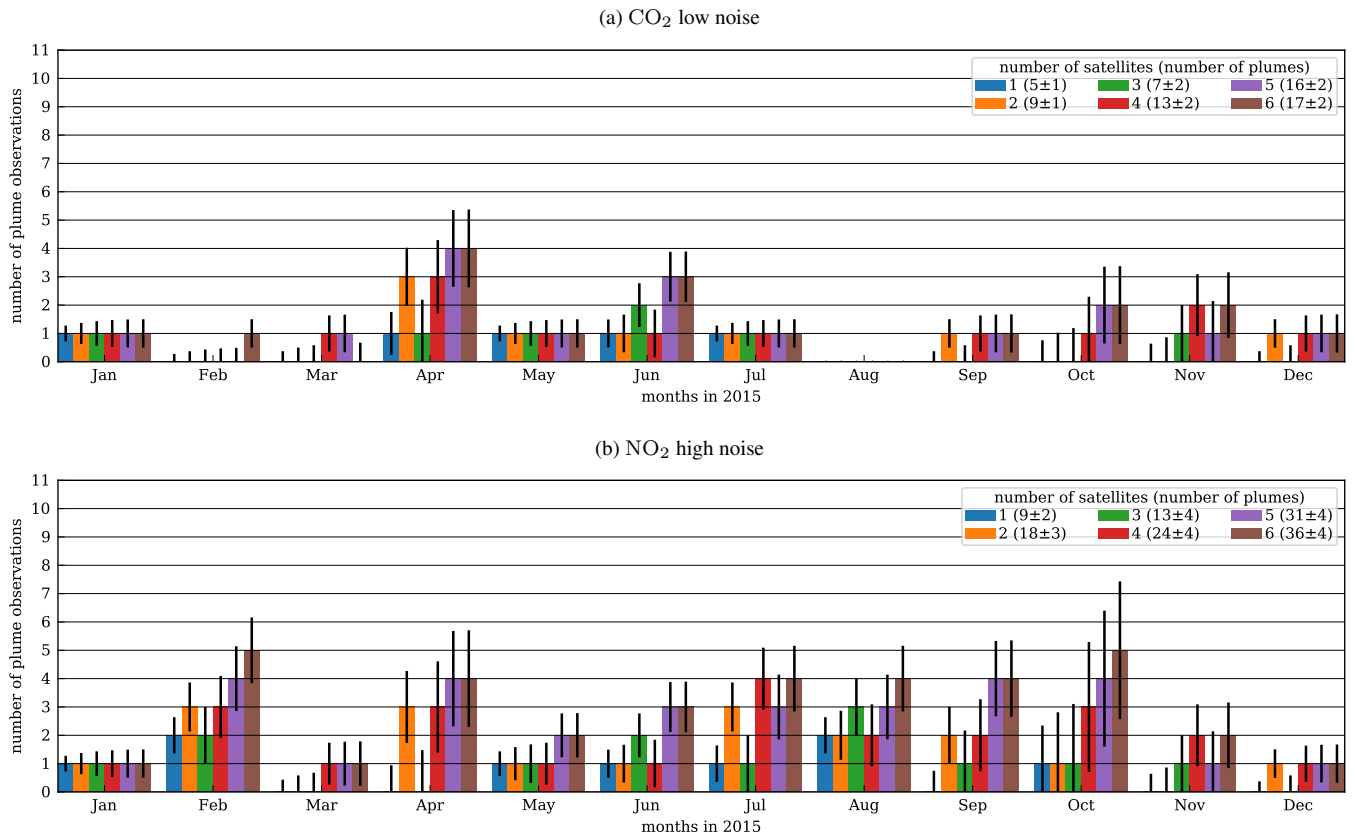
**Figure 11.** Number of detected plumes with at least 100 pixels (TP≥100 and PPV≥0.80) for one to six satellites with the (a) $CO_2$ low noise and (b) $NO_2$ high noise scenario. Error bars were estimated from all available satellites with different equator starting longitudes.

**Table 8.** Number of plumes detected for Jänschwalde with six satellites. The number of plumes are provided for plumes with at least 10 detected $CO_2$ or $NO_2$ pixels and in addition with at least 10 $CO_2$ pixels (cloud cover <1%). In both cases, plumes were included where neighboring plumes were detected in addition to Jänschwalde, i.e. detection with a large number of false ~~positves~~positives (e.g. Fig. 13c). These plumes are also shown separately. The classification uncertainty is about ±5 plumes. The neighborhood size was set to $n_s = 5$.

| Instrument scenario | ≥ 10 detected pixels | Number of plumes with ≥ 10 $CO_2$ pixels | large number of false positives |
|---|---|---|---|
| $CO_2$ low noise | 44 | 42 | 7 |
| $CO_2$ medium noise | 42 | 40 | 6 |
| $CO_2$ high noise | 41 | 40 | 4 |
| $NO_2$ low noise | 90 | 68 | 38 |
| $NO_2$ high noise | 91 | 68 | 34 |

of an individual power plant is not known. Therefore, we applied a visual inspection to identify those plume detections which erroneously included neighboring plumes. Furthermore, we limit the analysis to Jänschwalde.
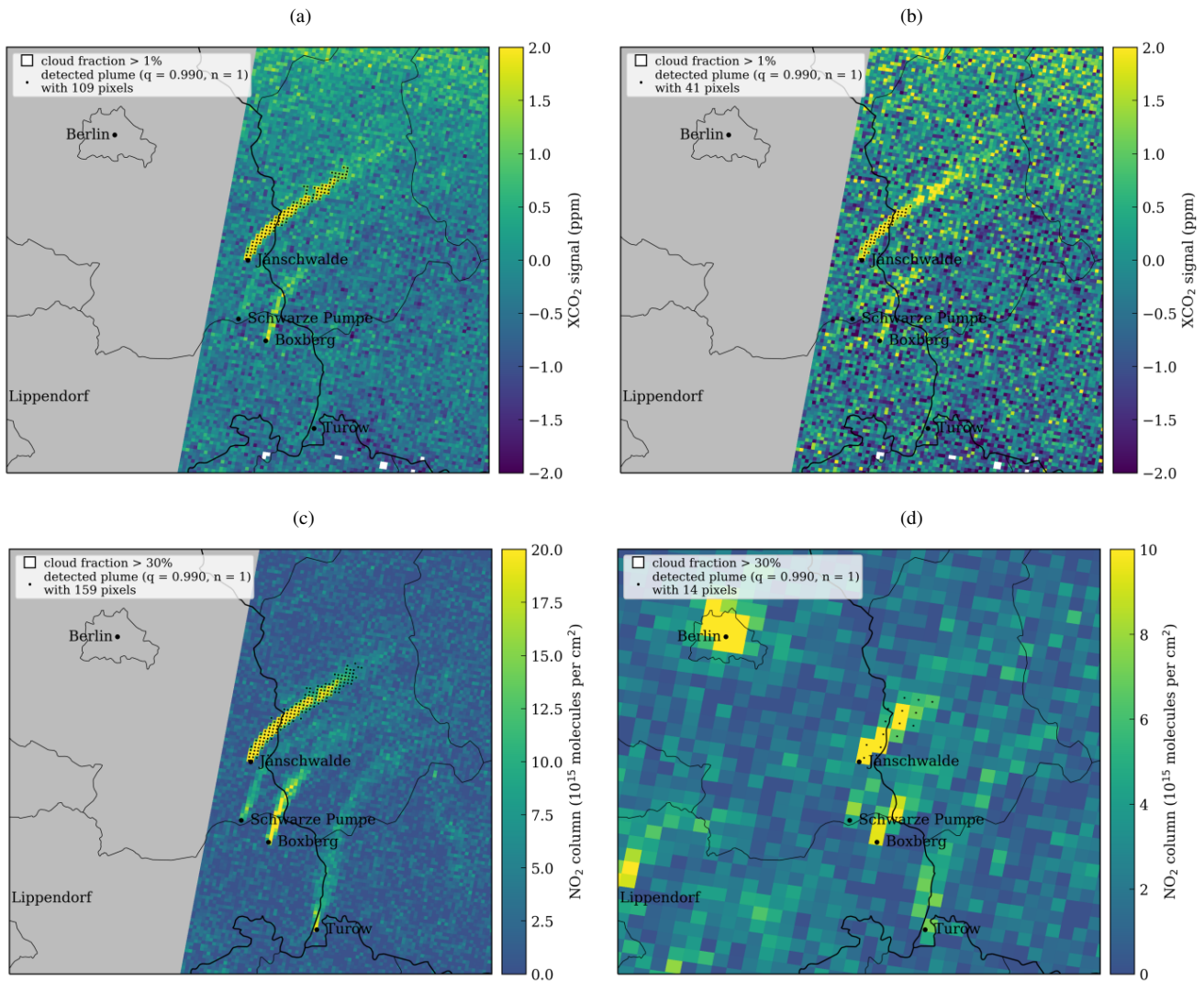
**Figure 12.** Example of plume detection for Jänschwalde power plant on 2 November 2015 using (a, b) the $CO_2$ instrument with $\sigma_{\text{VEG50}}$ of 0.5 and 1.0 ppm, (c) the $NO_2$ instrument with the high noise scenario and (d) the $NO_2$ instrument on Sentinel-5.

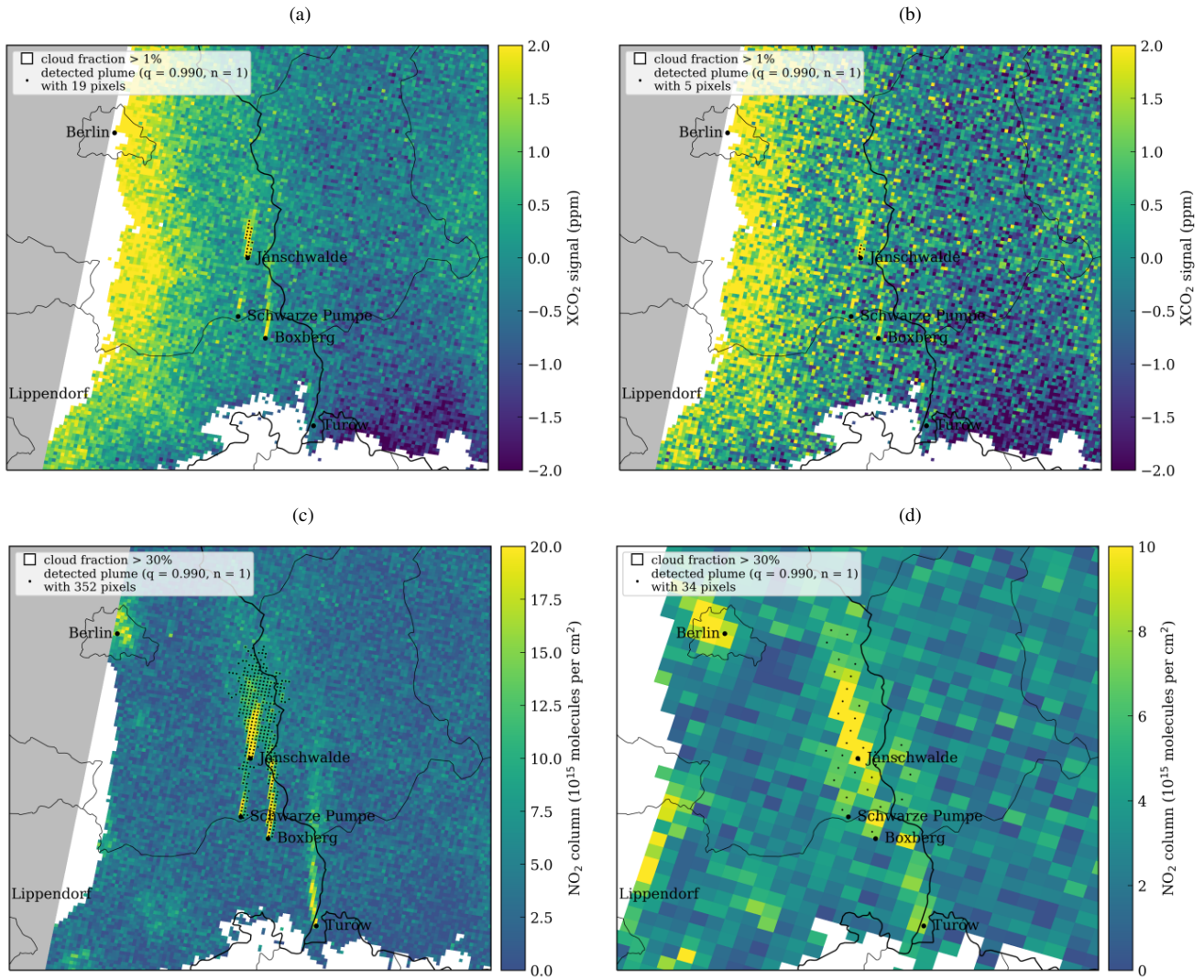**Figure 13.** Example of plume detection for Jänschwalde power plant on 17 February 2015 using (a, b) the $CO_2$ instrument with $\sigma_{\mathrm{VEG50}}$ of 0.5 and 1.0 ppm, (c) the $NO_2$ instrument with the high noise scenario and (d) the $NO_2$ instrument on Sentinel-5

As an example, Fig. 12 shows the successful detection of the $CO_2$ plume of Jänschwalde on 2 November 2015 by different instruments. Since $CO_2$ emissions of Jänschwalde are high with 33.3 Mt $CO_2$ yr$^{-1}$, the $XCO_2$ signal is very strong and can be detected well even with a high noise instrument ($\sigma_{VEG50}$ = 1.0 ppm). With low noise ($\sigma_{VEG50}$ = 0.5 ppm) the weaker plumes of Schwarze Pumpe and Boxberg are visible as well. The $NO_2$ instrument detects the four plumes in the region well. On this day also the Sentinel-5 $NO_2$ instrument successfully detects the plume of Jänschwalde and other point sources. Figure 13 presents a second, more challenging example for 17 February 2015. The $CO_2$ instrument successfully detects the plume with 0.5 ppm uncertainty, but with 1.0 ppm uncertainty, the number of detected pixels is likely too small to be useful for emission estimation. The reason for the low number of detected pixels in this case is the strong horizontal gradient in the $CO_2$ background. The $NO_2$ instrument detects the plume, but because the $NO_2$ plume of Jänschwalde overlaps with neighboring plumes, these plumes are erroneously assigned to Jänschwalde as well. At the coarser resolution of Sentinel-5 the plumes of the individual power plants can hardly be separated and, moreover, the time difference of two hours results in a plume location that is shifted with respect to the plume seen by the CO2M satellite.

Table 8 summarizes the results of the plume detection for Jänschwalde under the different instrument scenarios. It shows the number of detected plumes with at least 10 pixels and in addition, the number of plumes with at least 10 valid $CO_2$ pixels (cloud cover <1%). Note that in the case of the much narrower plumes from power plants, fewer pixels are required to form a useful plume. We classified detections that include large parts of the background as failed, but still counted detections that include neighboring plumes as successful (e.g. Fig. 13c), because they successfully identified the location of the plume. Since the classification is not always unambiguous, we assigned an uncertainty of about ±5 plumes at most.

In the year 2015, the number of detectable plumes with more than 10 pixels for a constellation of six satellites was between 40 and 45 for a $CO_2$ instrument with $\sigma_{VEG50}$ of 0.5, 0.7 and 1.0 ppm. At the same time, the $NO_2$ instrument detected about 90 plumes for the low and high noise scenario. When only plumes with more than 10 cloud-free $CO_2$ observations were considered, the number was reduced to about 70 plumes. For a smaller number of satellites, the number of detectable plumes would be correspondingly smaller. The $NO_2$ instrument detects more plumes because of its lower sensitivity to clouds, which makes it possible to trace the plume to the source even for partly cloudy scenes. On the other hand, the $NO_2$ instrument often detects overlapping plumes (e.g. Boxberg and Schwarze Pumpe), because the instrument is much more sensitive to small signals further away from the origin than the $CO_2$ instrument. The mean plume size was about 100 pixels for the low noise $CO_2$ instrument. The plumes detected with the high noise $CO_2$ instrument were about half the size. The $NO_2$ instruments detected a similar number of $CO_2$ pixels as the low noise $CO_2$ instrument, but when all detected pixels are counted the number of pixels doubles.

# 4 Discussion

## 4.1 Comparison with previous studies

In this study we investigated whether and how frequently the $CO_2$ plume of Berlin and power stations can be detected by different constellations of satellites using either $CO_2$ observations alone or in combination with observations of the co-emitted

trace gases CO and $NO_2$. To address the question, high-resolution $CO_2$, CO and $NO_2$ fields were simulated with the COSMO-GHG model for the year 2015 and used to generate synthetic $XCO_2$, CO and $NO_2$ satellite observations for Sentinel-5 and a constellation of CO2M satellites. Similar OSSEs studies were conducted by Pillai et al. (2016) and Broquet et al. (2018) for Berlin and Paris, respectively, as part of the LOGOFLUX study (Bacour et al., 2015). However, their simulations did not include $NO_2$ and CO fields. A fundamental difference is the realistic, i.e. not as a random noise, account for transport model errors in the present study, where the location of the plume is not taken from the model but detected in the satellite image using either $CO_2$ or $NO_2$ observations. For this reason, the focus of this paper is on the detectability of the plume, while Pillai et al. (2016) and Broquet et al. (2018) focused on the inversion, which we will describe in a follow-up publication.

Pillai et al. (2016) simulated $CO_2$ fields with the WRF-GHG model with $10 \times 10$ $km^2$ spatial resolution for the year 2008. The resolution was relatively low compared to the $1.1 \times 1.1$ $km^2$ resolution used in our study. They used $CO_2$ emissions from the EDGAR inventory (Version 4.1), which are more than twice as high as the emissions reported in the inventory of the city of Berlin as mentioned earlier. A consequence of the unrealistic high emissions are higher $CO_2$ signals (0.80 - 1.35 ppm) for Berlin than in our study (0.16 - 1.03 ppm, see Table 5). Note that the signal strength also depends on the spatial resolution, but our $XCO_2$ signals were computed for a local mean ($n_s = 37$, i.e. 148 $km^2$ spatial resolution) that is comparable to the model resolution used by Pillai et al. (2016). For Paris, Broquet et al. (2018) conducted simulations with the CHIMERE atmospheric transport model with 2 km spatial resolution. $CO_2$ emissions for the greater urban area were 40-50 Mt $CO_2$ $yr^{-1}$ that resulted in a $XCO_2$ signal of ~1 ppm, quite consistent with the plume signals reported here.

For Berlin, we estimated that 3 to 13 potentially useful $CO_2$ plumes (defined as plumes with at least 100 cloud-free pixels above a threshold of 0.05 ppm) would be observable, but not necessarily detectable, during one year by a single CO2M satellite with a 250-km wide swath. Pillai et al. (2016) identified 41 potentially useful orbits for estimating emissions with a 500-km wide swath. Although a direct comparison of these two numbers is difficult because of the different swath widths and the different definitions of "usefulness", we can still conclude that our study identified significantly fewer plumes than Pillai et al. (2016) even after halving their number to account for their wider swath.

Clouds have a strong impact on the number of cloud-free observations. A major difference between Pillai et al. (2016) and our study are the different approaches for masking cloudy observations. Pillai et al. (2016) used the MODIS cloud mask product (MOD35_L2) while we used cloud masks derived from cloud fractions simulated with the COSMO-GHG model. To compare these two approaches, we computed the number of cloud-free pixels using the MODIS cloud mask product (MOD35_L2), the MODIS cloud product (MOD06_L2) and COSMO-GHG simulations. The COSMO-GHG cloud fractions were spatially averaged over the MODIS pixels. The comparison shows that monthly fractions of cloud-free pixels agree well between masks computed from COSMO-GHG and MODIS cloud fractions for both a 1% and 30% cloud fraction, while the cloud-free pixels based on the MODIS cloud mask product are about twice as high (Figure S1 in the supplement). The larger fraction of cloud-free observations obtained from the MODIS cloud mask product is also consistent with a validation study

showing that the product is not very sensitive to optically thin clouds (Ackerman et al., 2008). The differences likely explain the different number of potentially useful orbits between Pillai et al. (2016) and our study. It also suggests that Pillai et al. (2016) overestimate the number of potentially useful orbits while our results are likely more accurate.

Further differences are to be expected because small plumes with less than 100 pixels were excluded in our case . The remaining difference is likely explained by their using and because of different meteorology, especially cloud cover and wind speed, for the different simulation periods. In addition, Pillai et al. (2016) used higher emissions and not considering did not consider vertical profiles of emissions, which results in stronger and correspondingly larger plumes(Brunner et al., 2019) . Furthermore, differences are to be expected due to different meteorology, especially cloud cover and wind speed, for the different simulation periods, which are less likely to be fully covered by clouds (Brunner et al., 2019).

Since the number of "useful" plumes observable by a satellite critically depends on its orbit (with one or two overpasses over Berlin per 11-day repeat cycle), the number of useful plumes per satellite may easily be overestimated if only an optimal orbit is considered. Comparing different equator starting longitudes as applied in this study reduces the sensitivity to a specific orbit selection.

## 4.2 Benefits of CO and $NO_2$ measurements

Out of 50 potentially useful plumes, many plumes were too weak to be easily seen by the $CO_2$ instrument. Plumes of Berlin with more than 100 detectable pixels could only be detected in about 12% of cloud-free cases with a high-noise instrument ($\sigma_{VEG50} = 1.0$ ppm) and about 32% with a low-noise instrument ($\sigma_{VEG50} = 0.5$ ppm). The success rate of 32% for an imperfect but still precise instrument ($\sigma_{VEG50} = 0.5$ ppm) would only allow for 2 to 3 favorable plume observations per year and satellite. These numbers illustrate the challenge and call for a larger constellation or a wider swath to increase the opportunities for plume detection and emission quantification, and for a $CO_2$ instrument with as low noise as possible.

Adding an $NO_2$ instrument greatly enhanced the opportunities for detecting the $CO_2$ plumes (68%-70% of cloud-free cases), since the $NO_2$ plumes largely overlap with the $CO_2$ plumes and since the signal-to-noise ratio is better for the $NO_2$ instrument. Furthermore, variability in the background is less important than for $CO_2$, and the $NO_2$ observations are less sensitive to clouds. Nevertheless, even with an $NO_2$ instrument, the number of detectable plumes per satellite remained small, of the order of 5 to 6 per year. A CO instrument with specifications similar to the CO instrument on Sentinel-5 had a smaller signal-to-noise ratio than the $CO_2$ instrument. Such an instrument would add little useful information over a developed region like Germany where combustion processes are well controlled and $CO:CO_2$ emission ratios correspondingly small. The question whether CO signals would be sufficiently high in other regions of the globe was outside the scope of this study.

The Sentinel-5 $NO_2$ instrument is well suited to detect the $NO_2$ plume of Berlin. However, the different overpass times of the CO2M satellite (11:30 local time) and Sentinel-5 (9:30 local time) frequently resulted in a significant spatial mismatch between the plumes, which reduced the number of matching plumes to 20% of the cloud-free cases. This is similar as for the $CO_2$ instrument with medium noise but three to four times lower than in the case of an $NO_2$ instrument placed directly on the CO2M satellite.

The detection of plumes from strong point sources like the power plant Jänschwalde was easier than the detection of city plumes, because point sources tend to have stronger and more confined $CO_2$ plumes for the same amount of emitted $CO_2$. In addition, the number of pixels required to map out such a plume was smaller, with only 10 detectable pixels being typically sufficient for identifying the main part of the plume. With a constellation of six satellites, about 40-45 plumes from Jänschwalde (33.3 Mt $CO_2$ $yr^{-1}$ emissions) with more than 10 detectable pixels could be observed per year even with a high-noise $CO_2$ instrument. This corresponds to 6 to 8 plumes per satellite and year, which is significantly better than for Berlin with the best instrument. The number of detectable plumes further increased by about 50% with the $NO_2$ instrument (about 70 plumes). Smaller point sources with emissions of about 10 Mt $CO_2$ $yr^{-1}$ (e.g. Lippendorf, Schwarze Pumpe and Turow) were sometimes detectable with a low-noise $CO_2$ instrument, but could also be detected with an $NO_2$ instrument.

Our study did not include systematic errors in the satellite observations ~~,~~ from aerosols, clouds and surface reflectance, which can result in spatial patterns resembling plume structures and therefore complicate plume detection. ~~Systematic~~ We therefore might overestimate the number of detectable plumes. Although systematic errors affect all satellite products, ~~but~~ the effect would be more severe for $XCO_2$ than for $NO_2$ due to the much smaller signal-to-noise ratios. How much aerosols enhance measurement uncertainties and correspondingly reduce the ability to detect plumes cannot be quantified here, but will be studied, for example, in a study on the use of aerosol information for estimating fossil fuel $CO_2$ emissions (AEROCARB) performed by a consortium led by SRON.

Our study could not simulate detailed $NO_X$ chemistry due to the high computational costs. We can therefore not rule out a larger mismatch between some $CO_2$ and $NO_2$ plumes than simulated here, for example, because of different $NO_2$ decay rates at different altitudes. However, Reuter et al. (2019) found no obvious mismatch when comparing co-located $CO_2$ and $NO_2$ plumes from OCO-2 and TROPOMI observations. A follow-up study for quantifying the effect of $NO_x$ chemistry would be certainly desirable.

Since the shape and extent of the plume can be imaged more accurately with an $NO_2$ instrument on the CO2M satellites, the $NO_2$ instrument can also be used to assess and correct transport simulations and improve these simulations through data assimilation.

## 4.3 Limitations of current plume detection algorithm

The main objective of this study was to compare the basic detectability of $CO_2$ plumes with $CO_2$, $NO_2$ and CO observations. Our plume detection algorithm ~~presented here worked well even with~~ was able to detect weak signals well below the single sounding precision, but tended to fail when the $CO_2$ or $NO_2$ field was complex, for example when several plumes from adjacent sources overlapped or when the background had a spatial gradient. These cases can be easily identified by a trained human as done in this study, but will have to be automatized for application at the global scale, for example by applying machine learning methods.

The algorithm assumed accurate knowledge of the mean and variance of the background, which ~~in a more realistic setup would have to be~~ were estimated directly from ~~the satellite observations~~ a simulated background tracer. It is possible that due to this optimistic assumption the number of detectable plumes was overestimates. On the other hand, mean and variance

were computed in a rather simple way from a large window centered on the source ($200\times200$ km$^2$ for Berlin). In the present algorithm, spatial gradients in the background field contributed to the variance of the background and thus reduced the ability for plume detection. However, such gradients, if sufficiently smooth, could potentially be accounted for in a more advanced algorithm through spatial interpolation of the background surrounding the plume.

5  ~~Because~~ When applied to real satellite observations the background and its variance could either be taken from a model or estimated directly from the satellite observations. Although atmospheric transport models have large uncertainties at the ~~scale level~~ of individual plumes, they could provide reasonable estimates of the $CO_2$ and $NO_2$ background. Likely the best option is to derive the background directly from the satellite data, which requires further development of the plume detection algorithm presented here. An improved algorithm could start with an a~~data-driven approach to plume detection is essential. Nevertheless,~~

10  ~~a well-considered inclusion of model data, climatologies and other measurements could help constrain and detect the location of the plume, for example, by estimating the a~~ priori ~~plume location based on wind direction and speed, by estimating mean and variance of the background , and by adjusting the size of the neighborhoods based on plume type (city or point source) and wind speed. Further improvements~~ estimate of plume location and background and would then be updated iteratively to improve both plume location and background. A model could be used here to determine a suitable a priori plume location

15  and background. An improved version of the algorithm presented ~~here have~~ in this paper has the potential for increasing the number of detectable plumes per satellite as well as the number of $CO_2$ pixels per plume.

## 5   Conclusions

In this paper the potential for detecting $CO_2$ plumes of the city of Berlin and neighboring power stations was investigated for the Copernicus anthropogenic $CO_2$ monitoring mission (CO2M), which is a proposed constellation of $CO_2$ satellites of the

20  European Copernicus program. Since the interference of biospheric $CO_2$ makes the identification of weak anthropogenic $CO_2$ plumes challenging, plumes were detected either from $CO_2$ observations or from observations of the co-emitted trace gases CO and $NO_2$. The study used high-resolution atmospheric transport simulations to create realistic $CO_2$, CO and $NO_2$ fields at $1\times1$ km$^2$ horizontal resolution to generate synthetic observations of $XCO_2$, CO and $NO_2$ for constellations of up to six CO2M satellites and one Sentinel-5 satellite.

25  For the city of Berlin about $50\pm5$ potentially "useful" $CO_2$ plumes were identified for the year 2015 for a constellation of six satellites, i.e. about eight plumes could be observed by a single CO2M satellite per year. This number is somewhat smaller than reported in earlier studies (Bacour et al., 2015; Pillai et al., 2016), ~~but it is consistent because small plumes with less than 100 pixels were excluded in our case~~ mainly because masking cloudy pixels based on the simulated cloud fields leads to less cloud-free observations than using the MODIS cloud mask product. Many of these fifty potentially observable plumes were

30  too weak to be easily detectable by the $CO_2$ instrument. Plumes with more than 100 detectable pixels could only be identified in 12% and 32% of cloud-free cases with a high-noise ($\sigma_{\mathrm{VEG50}} = 1.0$ ppm) and low-noise $CO_2$ instrument ($\sigma_{\mathrm{VEG50}} = 0.5$ ppm), respectively. A CO instrument with the uncertainty scenario used in this study had a signal-to-noise ratio that was lower than for the $CO_2$ instrument and was therefore not suitable for detecting $CO_2$ plumes. On the other hand, adding an $NO_2$

instrument significantly increased the number of detectable plumes (68%-70% of cloud-free cases), because $CO_2$ and $NO_2$ plumes generally overlapped well. The better performance of the $NO_2$ instrument was partly due to the higher signal-to-noise ratio and partly due to the lower sensitivity to clouds. The Sentinel-5 instrument was also well suited to detect the $NO_2$ plumes, but the different overpass times of the CO2M satellites (11:30 local time) and Sentinel-5 (9:30 local time) often resulted in a large spatial mismatch.

Strong point sources like the power plant Jänschwalde could be detected more easily with the $CO_2$ instrument (40-45 plumes), because the plumes were spatially more confined and the signals were stronger. The number of detectable plumes increased further with the $NO_2$ instrument by about 50% (about 70 plumes). The Sentinel-5 instrument could also detect the $CO_2$ plume, but could not always distinguish the plumes from neighboring power plants due to the lower spatial resolution. In addition, the spatial mismatch between CO2M and Sentinel-5 was large due to the 2-hour time difference between overpasses. Smaller point sources with emissions of about 10 Mt $CO_2$ $yr^{-1}$ were only detectable with a low-noise $CO_2$ instrument, but were in most cases readily detectable with an $NO_2$ instrument.

In this study, power plant plumes could be detected even with an $NO_2$ instrument with high noise. The power plants were equipped with wet scrubber technology for reducing $SO_2$ and $NO_x$ emissions but not with the latest available technology. Future updates using selective catalytic or non-catalytic reduction have the potential to further reduce $NO_x$ emissions by 20 to 50% (Lecomte et al., 2017), which would place higher requirements on the $NO_2$ instrument and would make the low noise scenario more beneficial.

This study demonstrates the huge benefit of adding an $NO_2$ instrument to a constellation of CO2M satellites for detecting city plumes and weaker point sources. The major advantages of the $NO_2$ instruments are the higher signal-to-noise ratio and the lower sensitivity to clouds. Therefore, adding an $NO_2$ instrument is highly recommended and the low-noise instrument is preferable for detecting also weaker ~~plumes~~and cleaner plumes in terms of $NO_2$ emissions. Furthermore, development of an advanced plume detection algorithms that can detect $CO_2$ plumes reliable will be essential for the application on an operational satellite.

*Author contributions.* GK designed the model experiments, conducted the simulations, and wrote the manuscript with input from all co-authors. GB followed the project as external advisor and contributed critical input to the manuscript. JM contributed the VPRM flux data

required for the simulations. VC ported the COSMO-GHG extension to GPUs. YM and AL accompanied the study as ESA project and technical officers, respectively, and provided critical inputs and reviews during all phases of the project. DB led the SMARTCARB project and contributed critical input to manuscript.

34

# References

Ackerman, S., Holz, R., Frey, R., Eloranta, E., Maddux, B., and McGill, M.: Cloud detection with MODIS. Part II: validation, Journal of Atmospheric and Oceanic Technology, 25, 1073–1086, https://doi.org/10.1175/2007JTECHA1053.1, 2008.

Ackerman, S., Menzel, P., F. R., and B., B.: MODIS Atmosphere L2 Cloud Mask Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, https://doi.org/10.5067/MODIS/MYD35_L2.061, 2017.

Agustí-Panareda, A., Massart, S., Chevallier, F., Boussetta, S., Balsamo, G., Beljaars, A., Ciais, P., Deutscher, N. M., Engelen, R., Jones, L., Kivi, R., Paris, J.-D., Peuch, V.-H., Sherlock, V., Vermeulen, A. T., Wennberg, P. O., and Wunch, D.: Forecasting global atmospheric $CO_2$, Atmos. Chem. Phys., 14, 11 959–11 983, https://doi.org/10.5194/acp-14-11959-2014, 2014.

AVISO GmbH and IE Leipzig: Erstellung der Berliner Emissionskataster Industrie, Gebäudeheizung, sonstiger Verkehr, Kleingewerbe, sonstige Quellen, Baustellen - Schlussbericht Juni 2016, Tech. rep., https://www.berlin.de/senuvk/umwelt/luftqualitaet/de/emissionen/download/Endbericht_Emissionkataster_2015.pdf, 2016.

Bacour, C., Boesch, H., Bovensmann, H., Breon, F.-M., Broquet, G., Buchwitz, M., Houweling, S., Klonecki, A., Krings, T., and Santaren, D.: LOGOFLUX 2 - CarbonSat Earth Explorer 8 Candidate Mission - Inverse Modelling and Mission Performance Study, Final report of ESA study contract nÂ°4000109818/14/NL/FF/lf, project led by NOVELTIS (France), Report, 2015.

Baldauf, M., Seifert, A., Forstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities, Monthly Weather Review, 139, 3887–3905, https://doi.org/10.1175/Mwr-D-10-05013.1, 2011.

Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO2 retrieval from space, Journal of Geophysical Research, 109, D04 311, 2004.

Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmospheric Measurement Techniques, 4, 1905–1928, https://doi.org/10.5194/amt-4-1905-2011, http://www.atmos-meas-tech.net/4/1905/2011/, 2011.

Bovensmann, H., Buchwitz, M., Burrows, J. P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., and Erzinger, J.: A remote sensing technique for global monitoring of power plant CO2 emissions from space and related applications, Atmos. Meas. Tech., 3, 781–811, https://doi.org/10.5194/amt-3-781-2010, http://www.atmos-meas-tech.net/3/781/2010/, 2010.

Broquet, G., Bréon, F.-M., Renault, E., Buchwitz, M., Reuter, M., Bovensmann, H., Chevallier, F., Wu, L., and Ciais, P.: The potential of satellite spectro-imagery for monitoring $CO_2$ emissions from large cities, Atmospheric Measurement Techniques, 11, 681–708, https://doi.org/10.5194/amt-11-681-2018, 2018.

Brunner, D., Kuhlmann, G., Marshal, J., Clément, V., Fuhrer, O., Broquet, G., Löscher, A., and Meijer, Y.: Accounting for the vertical distribution of emissions in atmospheric $CO_2$ simulations, Atmospheric Chemistry and Physics, 19, 4541–4559, https://doi.org/10.5194/acp-19-4541-2019, 2019.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y., and Löscher, A.: Carbon Monitoring Satellite (CarbonSat): assessment of atmospheric $CO_2$ and $CH_4$ retrieval errors by error parameterization, Atmos. Meas. Tech., 6, 3477–3500, https://doi.org/10.5194/amt-6-3477-2013, 2013.

C40 cities: 27 C40 cities have peaked their greenhouse gas emissions, retrieved from https://c40-production-images.s3.amazonaws.com/other_uploads/images/1923_Peaking_emissions_Media_Pack_Extended_version.original.pdf?1536847923, 2018.

Chimot, J., Breon, F.-M., Prunet, P., Vinuesa, J.-F., Camy-Peyret, C., Broquet, G., Chevallier, F., Renault, E., Houweling, S., Buchwitz, M., Bovensmann, H., Pillai, D., Reuter, M., Marshall, J., Brunner, D., Bergamaschi, P., Ciais, P., and Klonecki, A.: LOGOFLUX - CarbonSat Earth Explorer 8 Candidate Mission - Inverse Modelling and Mission Performance Study, Final report of ESA study contract n°40010537/12/NL/CO, project led by NOVELTIS (France), Report, 2013.

5  Ciais, P., Crisp, D., Gon, H. v. d., Engelen, R., Heimann, M., Janssens-Maenhout, G., Rayner, P., and Scholze, M.: Towards a European Operational Observing System to Monitor Fossil CO2 emissions - Final Report from the expert group. European Commission, Copernicus Climate Change Service, Report, 2015.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., Eldering, A., Fisher, B. M., Fu, D., Gunson, M. R., Mandrake, L., Osterman, G. B., Schwandner, F. M., Sun, K., Taylor,
10  T. E., Wennberg, P. O., and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmospheric Measurement Techniques, 10, 59–81, https://doi.org/10.5194/amt-10-59-2017, 2017.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H. M., and Sweeney, C.: A climate-scale satellite record for carbon monoxide: the MOPITT Version 7 product, Atmospheric Measurement Techniques, 10, 2533–2555, https://doi.org/10.5194/amt-10-2533-2017, 2017.

15  Düring, I., Bächlin, W., Ketzel, M., Baum, A., Friedrich, U., and Wurzler, S.: A new simplified $NO/NO_2$ conversion model under considera-tion of direct $NO_2$-emissions, Meteorologische Zeitschrift, 20, 67–73, https://doi.org/10.1127/0941-2948/2011/0491, http://dx.doi.org/10.1127/0941-2948/2011/0491, 2011.

ESA: Report for mission selection: CarbonSat, ESA SP-1330/1 (2 volume series), Report, 2015.

Fioletov, V. E., McLinden, C. A., Krotkov, N., and Li, C.: Lifetimes and emissions of SO2 from point sources estimated from OMI, Geo-
20  physical Research Letters, 42, 1969–1976, https://doi.org/10.1002/2015GL063148, 2015.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geoscientific Model Development, 8, 975–1003, https://doi.org/10.5194/gmd-8-975-2015, 2015.

25  Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., and Courrèges-Lacoste, G. B.: Requirements for the GMES At-mosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p, Remote Sensing of Environment, 120, 58 – 69, https://doi.org/10.1016/j.rse.2012.01.023, the Sentinel Missions - New Opportunities for Science, 2012.

Krings, T., Gerilowski, K., Buchwitz, M., Hartmann, J., Sachs, T., Erzinger, J., Burrows, J. P., and Bovensmann, H.: Quantification of methane emission rates from coal mine ventilation shafts using airborne remote sensing data, Atmos. Meas. Tech., 6, 151–166,
30  https://doi.org/10.5194/amt-6-151-2013, https://www.atmos-meas-tech.net/6/151/2013/, 2013.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: TNO-MACC_II emission inventory; a multi-year (2003-2009) consistent high-resolution European emission inventory for air quality modelling, Atmos. Chem. Phys., 14, 10 963–10 976, https://doi.org/10.5194/acp-14-10963-2014, http://www.atmos-chem-phys.net/14/10963/2014/, 2014.

Kuhlmann, G., Clément, V., Marshall, J., Fuhrer, O., Broquet, G., Schnadt-Poberaj, C., Löscher, A., Meijer, Y., and Brunner, D.: SMART-
35  CARB - Use of Satellite Measurements of Auxiliary Reactive Trace Gases for Fossil Fuel Carbon Dioxide Emission Estimation, Final report of ESA study contract n°4000119599/16/NL/FF/mg, Tech. rep., Empa, Swiss Federal Laboratories for Materials Science and Tech-nology, Dübendorf, Switzerland, https://www.empa.ch/documents/56101/617885/FR_Smartcarb_final_Jan2019.pdf, 2019.

Lecomte, T., de la Fuente, J. F. F., Neuwahl, F., Canova, M., Pinasseau, A., Jankov, I., Brinkmann, T., Roudier, S., and Sancho, L. D.: Best Available Techniques (BAT) Reference Document for Large Combustion Plants - Industrial Emissions Directive 2010/75/EU Integrated Pollution Prevention and control, Tech. rep., 2017.

Liu, Y., Gruber, N., and Brunner, D.: Spatiotemporal patterns of the fossil-fuel CO2 signal in central Europe: Results from a high-resolution atmospheric transport model, Atmos. Chem. Phys., 2017, 14 145–14 169, https://doi.org/10.5194/acp-17-14145-2017, 2017.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochemical Cycles, 22, https://doi.org/10.1029/2006GB002735, 2008.

MOPITT Algorithm Development Team: Measurements of Pollution in the Troposphere (MOPITT) Version 7 Product User's Guide, https://www2.acom.ucar.edu/sites/default/files/mopitt/v7_users_guide_201707.pdf, 2017.

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., and Crisp, D.: Quantifying $CO_2$ Emissions From Individual Power Plants From Space, Geophysical Research Letters, 44, 10,045–10,053, https://doi.org/10.1002/2017GL074702, 2017.

Oney, B., Henne, S., Gruber, N., Leuenberger, M., Bamberger, I., Eugster, W., and Brunner, D.: The CarboCount CH sites: characterization of a dense greenhouse gas observation network, Atmospheric Chemistry and Physics, 15, 11 147–11 164, https://doi.org/10.5194/acp-15-11147-2015, http://www.atmos-chem-phys.net/15/11147/2015/, 2015.

Pillai, D., Buchwitz, M., Gerbig, C., Koch, T., Reuter, M., Bovensmann, H., Marshall, J., and Burrows, J. P.: Tracking city CO2 emissions from space using a high-resolution inverse modelling approach: a case study for Berlin, Germany, Atmos. Chem. Phys., 16, 9591–9610, https://doi.org/10.5194/acp-16-9591-2016, http://www.atmos-chem-phys.net/16/9591/2016/, 2016.

Pinty, B., Janssens-Maenhout, G., Dowell, M., Zunker, H., Brunhe, T., Ciais, P., Dee, D., van der Gon, H. D., Dolman, H., Drinkwater, M., Engelen, R., Heimann, M., Holmlund, K., Husband, R., Kentarchos, A., Meijer, Y., Palmer, P., and Scholz, M.: An Operational Anthropogenic CO2 Emissions Monitoring & Verification Support capacity - Baseline Requirements, Model Components and Functional Architecture, Report, https://doi.org/10.2760/39384, 2018.

Platnick, S., Ackerman, S., King, M., Wind, G., Meyer, K., Menzel, P., Frey, R., Holz, R., Baum, B., and Yang, P.: MODIS atmosphere L2 cloud product (06_L2), NASA MODIS Adaptive Processing System, Goddard Space Flight Center, https://doi.org/10.5067/MODIS/MYD06_L2.061, 2017.

Reuter, M., Buchwitz, M., Hilboll, A., Richter, A., Schneising, O., Hilker, M., Heymann, J., Bovensmann, H., and Burrows, J.: Decreasing emissions of NO x relative to CO 2 in East Asia inferred from satellite observations, Nature Geoscience, 7, 792, https://doi.org/10.1038/ngeo2257, 2014.

Reuter, M., Buchwitz, M., Schneising, O., Krautwurst, S., O'Dell, C. W., Richter, A., Bovensmann, H., and Burrows, J. P.: Towards monitoring localized $CO_2$ emissions from space: co-located regional $CO_2$ and $NO_2$ enhancements observed by the OCO-2 and S5P satellites, Atmospheric Chemistry and Physics, 19, 9371–9383, https://doi.org/10.5194/acp-19-9371-2019, 2019.

Schaaf, C. and Wang, Z.: MCD43A3 MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global - 500m V006 [Data set]. NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MCD43A3.006, 2015.

Schaub, D., Brunner, D., Boersma, K. F., Keller, J., Folini, D., Buchmann, B., Berresheim, H., and Staehelin, J.: SCIAMACHY tropospheric NO2 over Switzerland: estimates of NOx lifetimes and impact of the complex Alpine topography on the retrieval, Atmospheric Chemistry and Physics, 7, 5971–5987, https://doi.org/10.5194/acp-7-5971-2007, http://www.atmos-chem-phys.net/7/5971/2007/, 2007.

Sierk, B., Bézy, J.-L., Löscher, A., and Meijer, Y.: The European $CO_2$ Monitoring Mission: observing anthropogenic greenhouse gas emissions from space, https://doi.org/10.1117/12.2535941, 2019.

Taylor, T. E., O'Dell, C. W., Frankenberg, C., Partain, P. T., Cronk, H. Q., Savtchenko, A., Nelson, R. R., Rosenthal, E. J., Chang, A. Y., Fisher, B., Osterman, G. B., Pollock, R. H., Crisp, D., Eldering, A., and Gunson, M. R.: Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: validation against collocated MODIS and CALIOP data, Atmos. Meas. Tech., 9, 973–989, https://doi.org/10.5194/amt-9-973-2016, https://www.atmos-meas-tech.net/9/973/2016/, 2016.

5  Ting, K. M.: Confusion Matrix, pp. 209–209, Springer US, Boston, MA, https://doi.org/10.1007/978-0-387-30164-8_157, 2010.

UNFCCC: Paris Agreement, FCCC/CP/2015/L.9/Rev1, retrieved from http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf, 2015.

v. Storch, H. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambride University Press, 2003.

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, Atmospheric Measurement Techniques, 11, 5673–5686,

10  https://doi.org/10.5194/amt-11-5673-2018, 2018.

Velazco, V. A., Buchwitz, M., Bovensmann, H., Reuter, M., Schneising, O., Heymann, J., Krings, T., Gerilowski, K., and Burrows, J. P.: Towards space based verification of CO2 emissions from strong localized sources: fossil fuel power plant emissions as seen by a Carbon-Sat constellation, Atmos. Meas. Tech., 4, 2809–2822, https://doi.org/10.5194/amt-4-2809-2011, http://www.atmos-meas-tech.net/4/2809/2011/, 2011.

15  Wenig, M. O., Cede, A. M., Bucsela, E. J., Celarier, E. A., Boersma, K. F., Veefkind, J. P., Brinksma, E. J., Gleason, J. F., and Herman, J. R.: Validation of OMI tropospheric NO2 column densities using direct-Sun mode Brewer measurements at NASA Goddard Space Flight Center, Journal of Geophysical Research, 113, D16S45, 2008.