Author's response:

We thank the reviewers for their detailed comments and suggestions. We are confident that we addressed all issues raised by the reviewers appropriately and thus significantly improved the quality of the manuscript. Before discussing all comments in detail, we would like to clarify some important aspects that we treated during the review process.

A main concern for both reviewers was the averaging time of the measured NO_2 concentration. We reported and used 3 s average values for our statistical modeling. The reviewers suggest that this time should be chosen equal to the flushing time of the measurement cell which is about 10 s. However, it is important to realize that the flushing of a fully mixed reactor (here the measurement cell) corresponds to a low pass filter. This does not imply that averaging must be done at the response time of this setup. Shorter averaging increases the noise of the individual measurements, while longer averaging leads to loss of information (see e.g. Nyquist criterion). Within that context, we still consider 3 s averages as a good compromise and did not change the time resolution.

However, after careful reconsideration, we came to the conclusion that adding a time shift to the tram's GPS position corresponding to the instrument response time is not appropriate either. This is because the instrument response is instantaneous (except for a negligible delay caused by the air travelling from the inlet to the measurement cell) even though filtered by a low pass filter. Therefore, we removed that time shift and recalculated the statistical modelling. We exchanged the corresponding figures and updated the results tables. This had no significant impact on the results and their interpretation.

Finally, we changed figures 1 and 3 according to the reviewers' comments and added details to the description of the instrument and measurement technique. We also added information about the CLD instruments that were used for comparison and made clear that they are operated in agreement with the European Standard EN14211, which is mandatory for regulatory measurements and assures comparability and quality of the CLD measurements.

In the following we address all comments of both reviewers in detail.

Reviewer #2

I) General comments:

Comment 1) "The need of mobile measurements is well introduced but an overview of available instruments is missing (commercial or laboratory prototypes). Here CLDs and electrochemical sensors are mentioned and references are given for instruments previously developed by the authors in the MIR. However there is no mention to the many different spectroscopy techniques used in the MIR as well as in the blue spectral region with ECDL or LED. Please compare the performances of the spectrometer developed with these previous works that have to be cited in this paper."

Response In the introduction we added further references and included previous works (CAPS, ECDL, LED) as suggested.

Modification *"Laser spectroscopy, in contrast to CLD and electrochemical sensors, determines the concentration of* NO_2 *directly and with sampling rates of 1 Hz and higher. Previous studies have shown the applicability of laser diodes in the blue spectral region to measuring* NO_2 *by various cavity enhanced techniques (Courtillot et al., 2006; Fuchs et al., 2009; Kebabian et al., 2005; Osthoff et al., 2006). With higher absorption cross sections than in the visible the mid-infrared region is often chosen for high precision trace gas monitoring of* NO_2*. Here, for instance photoacoustic spectroscopy (Pushkarsky et al., 2006), quartz-enhanced photoacoustic spectroscopy (Patimisco et al., 2014) and direct absorption spectroscopy (McManus et al., 2015) were applied."*

C 2) "P2,L28" The portable instrument was specifically built for this campaign". Figure 1 is just a scheme that could fit to all multipass spectrometer. The reader expects a detailed description of the technique (iCW mode) and the set-up to understand how the good performances reported in the following are reached. A spectra recorded by the instrument should be shown. Please, detail the "Quantum cascade laser spectrometer" section in this way (see detailed comments in the following)."

R We modified the description of the instrument and figure 1 accordingly to make the key elements that contributed to the performance more clear.

An absorption spectrum recorded with the instrument was added as figure 1(c).

M The modifications can be found with the responses to the detailed comments.

C 3) "The statistical model is based on average NO2 measurements during 3s (P6,L18- P9,L15). Note that the response time of the instrument is longer: 9.6s. Averaging at this time scale should be performed prior to analysis."

R We reconsidered the data analysis procedure. The instrument's response to a change in the ambient NO_2 concentration is instantaneous except for the negligible travel time of an air parcel from the inlet to the multipath cell. The mixing of air inside the multipath cell effects that the measured NO_2 concentration does not immediately correspond to the NO_2 concentration pumped into the cell. This effect was quantified by these 9.6 seconds. The instrument inherently smooths temporal variations in NO_2 concentration acting as a low pass filter.

We keep the original three seconds sampling rate but leave out the shift of the measurements by 9.6 seconds. Temporal shifting of the measurements is not appropriate. Additional averaging of the data to 9 seconds would further reduce information of the spatial variability of the NO_2 concentration field without resolving the smoothing effect inherent to the measurement instrument.

We recomputed the statistical models with data processed according to the adapted procedure. Results are reported based on these computations.

Shorter sampling intervals than 3 seconds or alternative sampling schemes (e.g. 1s every 3s) may further improve data quality and the assignment of a NO_2 concentration to a position but were not implemented in this study. We think that 3 second average values are adequate to the chosen complexity of our statistical models.

C 4) "NO2 concentration is expected to be sensitive to the solar activity, NO, and O3 concentrations, and to ambient temperature and humidity. Variations of NO2 measurements should be studied according to these different parameters. It is hard to believe that all these parameter effects are removed by the offset correction."

R This is a misunderstanding: Our offset correction does not "remove" any of the mentioned influences on the concentration, but is solely applied to correct for instrumental drifts that appear as a shift in the measured zero. The reviewer is of course correct that ambient NO_2 concentrations depend on the mentioned factors. However, the goal here is to determine the ambient NO_2 concentration and not to investigate the processes leading to the formation or transformation of NO_2. Therefore the task is to minimize sampling artefacts, e.g. through formation of NO_2 in the inlet from reaction of NO with O_3, which is inherent for every measurement using a closed measurement cell. The residence time of the air in the sampling line and the flushing time in the measurement cell have been kept as short as possible, so that the sampling artefacts are minimized. The QCL measurements are corrected to be consistent with the reference CLD measurements. During mobile deployment, the repeated comparison of the QCL measurements and NO_2 concentrations measured at the fixed air quality monitoring sites support that the NO_2 sampling artefacts (or losses) are either small or constant on relative terms.

C 5) "QCLAS measurements is compared to CLDs measurements taken as reference values. Specification of the CLDs are expected : sensitivity, absolute accuracy, intercomparison of the CLDs..."

R CLD is the standard technique for air-quality monitoring of NO_x. The instruments are operated according to the European standard EN 14211, including ongoing quality control procedures such as regular calibrations and maintenance by the monitoring networks including intercomparison of the CLDs (Reference: "Technischer Bericht zum Nationalen Beobachtungsnetz für Luftfremdstoffe" (https://www.empa.ch/documents/56101/246436/Technischer+Bericht+2016/0bc321a3-f489-4f20-bcda-a323fbc4ca8a), 2016).

We added detailed specifications of the performance of the CLDs as determined within the Swiss Air Pollution Monitoring Network NABEL.

M *"The CLD has a limit of detection of 0.1 ppb and at 30 ppb a total uncertainty of 1.25 ppb (NABEL, 2016)."*

Detailed comments

C "**a)** In the abstract and the introduction: please mentioned the technique used and the wavenumber of the laser."

R Done as suggested.

M *"The NO_2 concentration is measured by direct absorption spectroscopy at 1600 cm^{-1}."*

C "**b)** Introduction:

1) This work is dedicated to urban monitoring, please specify the NO2 concentration range to clarify the instrument performances required in term of detection limit.

2) Insert references to previous NO2 spectrometers (see General comments)

3) Please gives accurate numbers concerning instrument performances:

P1, L30 "accurate"?

P2, L9: "not fast enough"?

P2 L11 "not short enough" P2 L13 "relatively high"

P2 L16: "high sampling rates"

R The comments above were implemented as suggested.

C "**c)** 1.1 QCL spectrometer :"

Ten comments concerning the QCL spectrometer were made. In the following we will address these comments individually.

C "1) The iCW operation was published in Fischer 2014, reporting on a different instrument that operates at 7.7µm. In this paper, a short description of this technique is expected and a detailed description of the optical set-up at 6.25µm as well. I could not find mention to an optical etalon : How is obtained the frequency linearization ?"

R We thank the reviewer for pointing this out. We modified the respective sections to include details about icw driving, the optical set-up and the frequency linearization with a 2'' Ge etalon.

M *"In iCW mode the drive current of the QCL is dropped to zero between individual pulses leading to rapid heating and therefore frequency tuning during each pulse. The pulses are generated by discharging a capacitor over the QCL. A frequency tuning of >1 cm^{-1} was achieved (see Figure 1(c). By shaping the current ramp with different RC-elements, close to linear tuning could be achieved. See (Fischer et al., 2014) for details.*

…For frequency calibration a 5.08 cm Ge-etalon can be inserted between the laser and the first mirror."

C "2) Please show a recorded spectra around 1600 cm-1. How many NO2 lines do you monitor ? Is the spectra large enough to monitor water ?"

R A typical experimental spectrum of NO_2 and water absorption lines is included (Figure 1(c)). In fact, the water absorption line was always fitted along the NO_2 and used as a frequency reference for the NO_2 absorption and as a criterion to determine if the spectrum is fitted correctly.

M A spectrum was added as Figure 1(c).

C "3) In figure 1 : the detailed optical scheme would be more useful than the schematic content of the instrument"

R/M We replaced figures 1a and 1b with a detailed optical scheme and a photo of the instrument without cover (also following the suggestion of Reviewer #3).

C "4) P3, L17 : please specify that the 100µm orifice is placed at the inlet of the cellule and give the corresponding flow value."

R Done as suggested.

M *"...while the air flux is limited to 180 sccm by a 100 µm diameter orifice placed at the gas inlet."*

C "5)P3 L17-19 : The response time is a main concern. It is measured to be equal to 9.6s by switching from NO2 free air to NO2 sample. Is this really consistent with the complete exchange time of the cell according to the flow and pressure values?"

R The flushing time can also be estimated from the flow through the orifice which is about 180 sccm (=3.04 mbar·l/s). With a volume of 0.3 l and a pressure of 100 mbar the resulting flushing time would be about 10 s which is in good agreement with the measurement.

C "6) P3L22 : what is the acquisition rate ?"

R Spectra are generated at a rate of 2 kHz. However, due to the data transfer limitations of the DAQ system, only 50 % duty-cycle acquisition is achieved, i.e. 1000 spectra are averaged within 1 s. We will make this clear in the text.

M *"After acquisition of 1000 spectra at a rate of 2 kHz the data is transferred via a USB 2.0 port to a nano PC (Zero pro, Xi3 corporation) where the spectra are averaged and analyzed by a custom-written LabVIEW program."*

C "7)P3, L23-25: Concerning the fitting procedure : Please indicates which parameters are fixed (such as P probably) or left variable. The instrument not being temperature stabilized, what about the temperature parameter ? Is the path length deduced fixed?"

R We agree that this should be mentioned. The pressure in the cell is constant and could therefore be fixed. The temperature of the cell is measured by a thermistor and the actual temperature is fed back to the fitting routine. As the cell is a monolithic ring the path length is constant and therefore fixed in the fitting routine as well.

M *"Pressure and path length are considered constant, while the temperature of the cell is measured with a thermistor and fed back to the fitting program."*

C "8) Water absorption has strong absorption lines around 1600cm-1, and its concentration is changing during tram measurements. How is that taken into account ? Do you include a fit of water line to monitor it ? is it included in the base line?"

R Indeed, water has strong absorptions near our target NO_2 line. The pressure broadened wing of this line is included in the baseline of the fit therefore changes in water concentration do not change the result of the fit. Furthermore, a small HDO line is fitted along the NO_2 to monitor the water concentration and to serve as a frequency reference when the NO_2 concentration is close to our detection limit. The fitted lines can be seen in the spectrum that was added in response to *General comment 2*.

C "9) P3,L29: please make explicit what you call "zero-point offset"."

R We will rephrase to make clear what is meant by "zero point offset".

M *"Second, the instrument repeatedly determines the zero point offset, i.e. the instrument signal at zero NO_2, by measuring filtered NO_2-free air for 2 minutes every 20 minutes."*

C "10) The dynamic range of the instrument should be given somewhere in the paper"

R We agree and added information about the dynamic range of the instrument. The dynamic range goes from from 1 ppb to several ppm in concentration.

M *"The chosen NO_2 absorption feature at 1599.9 cm^{-1} offers a large dynamic range from 0 to several ppm which is suitable for ambient concentration measurements in cities."*

C "**d)**2.3. NO2 concentration measurements at fixed sites

Indicates that it is performed by CLD. Please give the specifications of these instruments (see general comments)."

R/M We will make clear that the instruments at the fixed sites are CLDs and give their specifications.

C "**e)** P5 L17 : " NO2 concentrations encountered in Zurich are moderate compared to other cities." Please specify here the typical range."

R We added information about the annual mean concentrations to the manuscript. Additionally, figure 6 shows a 24 h time series of NO_2 concentrations.

M *"The reported annual mean NO_2 concentrations in 2015 for the AQM sites were: HEU: 18 µg/m^3, ZUE: 31 µg/m^3, STA: 33 µg/m^3, SCH: 45 µg/m^3, SWD: 47 µg/m^3, RGS: 50 µg/m^3* (OSTLUFT, 2015)."

C "**f)** 2.5 Statistical modelling

1)P6, L18 : NO2 concentrations should be averaged to 9.6s, the response time (see general comments)"

R See our response to *general comment 3)*.

C "2)Please provide the reference (Breinam1984) concerning the "regression tree approach at first mention: P6, L32"

R We provided the reference as suggested.

C "**g)** 3.1. Instrument performance and stability"

Comments concerning the Instrument performance and stability were made. In the following we will address these comments individually.

C "1) The Allan deviation is expected to increase at long integration time due to technical drifts or ambient parameters change. Can the author provide a longer set of measurements to show when it appears ?

R We agree. The Allan-Werle plot (Figure 3) was updated using a longer time series. Now the onset of the drift after an integration time of about 400 seconds can be observed.

C "2) In figure 3, in the upper panel can you confirm that Dark-Noise measurements are the upper one or is there a color inversion ? I would understand an offset of about 1.5ppb is the measured in" NO2 free air" rather than Dark-Noise, from a non perfect filtering or other effects."

R There is no color inversion. Both traces should be around zero. It is rather an offset that was added to the zero-air measurement (red trace) for clarity.

M *"The dark-noise trace is offset by 1.5 ppb from zero for clarity."*

C "3)P8, L1 : Concerning the "measurement strategy" I understand how the drift is corrected but this process relies on the zero NO2 gas production in-situ. Did you test that there is no NO2 trace in this gas ? I would expect some NO2 trace at the ppb level (according to fig 1). And the residual concentration is expected to depend on NO2 concentration in ambient air. Furthermore, the sensitivity to water concentration should be investigated . Can you provide some details ?"

R Indeed, our drift correction fully relies on the quality of the in-situ zero-air generation. The Purafill chemisorbant is specified to have a minimal initial removal efficiency of 99.5%. That would leave a maximal trace of 0.1 ppb at the average ambient concentration in Zurich, which is 20 ppb. For a trace of 1 ppb an ambient concentration of 200 ppb would be necessary. The chemisorbant is specified to perform from 10-95% relative humidity and between -20 and 51°C. Laboratory tests confirmed that there is no detectable NO_2 after the filter.

C 3b"Another aspect should be detailed : how is calibrated the correction, on which reference measurement ? Did you performed a referenced measurement prior in-situ measurements ? Did you repeat it during the measurement campaign?"

R The linear correction to the NO_2 concentration values obtained by the QCLAS instrument was found by comparison with a calibrated CLD from the fixed air quality monitoring station in Dübendorf. The instrument measured in parallel with the CLD for several days. Figure 4 shows the agreement of the CLD and QCLAS measurements (slope 1) after the correction was applied. Such a calibration was not repeated during the measurement campaign. However, as detailed in the paper, we compared the QCLAS measurements in the vicinity of air-quality monitoring stations to the values reported by the stations (Figures 7a and b). The comparison showed very good agreement (slope 1.03) at the site STA, where a relatively small temporal and spatial variability is expected. Based on these measurements we are confident that the calibration did not change during the measurement campaign.

C "4) Allan deviation only demonstrate the stability of the instrument but not the accuracy. Caption of figure 3 is to be modified and 2 P7, L24-25 : "precision" should be replaced by "stability""

R Here must be a misunderstanding, because on P7, L24-25 we stated: "The best precision of 30 ppt is reached after 100 s averaging while the 1 s precision is about 300 ppt." and in the caption of figure 3: "Determination of the instrument precision and stability: Time series of zero-air (black) and detector dark-noise (red) and the associated Allan-deviation plots." There is no mention of "accuracy" here.

In addition, we do not agree that "precision should be replaced with stability" as the integration time of the Allan plot (x-axis) is a measure for stability, while the best precision can be seen from the corresponding Allan-deviation (y-axis).

C "5) P8, L8-14 : I would expect the correction factor to change in time due to passivation of the surfaces or water interferences : the calibration performed in June is probably different during winter. Did you test this point ?

R We agree that the correctness of the calibration factor has to be verified. For this reason we compared the QCLAS measurements to measurements at fixed air-quality monitoring sites when the instrument was in their vicinity which yielded good agreement (see also response to *comment 3b)*. Therefore we are confident that the calibration factor did not change significantly. Water interferences are excluded by the line selection as detailed in response to *comment c) 8)*.

C "6)The correction factor of 1.2 is applied. But how was the instrument firstly calibrated?"

R Please see response to comment g)3b).

C "7)P9L2: Please specify the absolute accuracy of the CLD to justify the 1ppb accuracy of the instrument."

R The CLD has a total uncertainty of 1.25 ppb at 30 ppb concentration (Reference: "Technischer Bericht zum Nationalen Beobachtungsnetz für Luftfremdstoffe" (https://www.empa.ch/documents/56101/246436/Technischer+Bericht+2016/0bc321a3-f489-4f20-bcda-a323fbc4ca8a), 2016). The accuracy of the QCLAS instrument has to be considered relative to the CLD which is the standard instrumentation for ambient air quality monitoring (see European Norm EN 14211). This will be changed in the manuscript to make it clear.

M *"The CLD has a limit of detection of 0.1 ppb and at 30 ppb a total uncertainty of 1.25 ppb (NABEL, 2016)."*

"...therefore we conclude that the instrument accuracy is about 1 ppb with <u>respect to the CLD</u> and that the assumption of a constant or negligible relative loss of NO_2 was correct."

C "8) P9L15-19 : as mentioned previously, measurements should be averaged to 9.6s resulting in a longer route segment."

R For the reasons stated above we are convinced that using 3 second averages is appropriate.

C "**h)** 3.2 Intra-urban and temporal variation in NO 2 concentration"

Five comments concerning Intra-urban and temporal variation in NO_2 concentration were made. In the following we will address these comments individually.

C "1) Fig 6 : in the caption please clarify which AQM station (STA or SCH) is close to which terminal stations (Seebach or Triemli)"

R We indicate the passing of the tram at the terminal stations primarily to show that the tram is in regular operation on October 23, 2015. A direct comparison of QCLAS measurements and measurements from the AQM sites is presented in Figure 7. Travel time of the tram service number 14: Seebach → AQM STA (+ ~16 min) → AQM SCH (+ ~16 min) → Triemli (+ ~7 min) [total: ~39 min].

C "2)P11L7 : again measurements should be averaged to 9.6s."

R See our answer above.

C "3)P11,L15-17: I don't understand why "this analysis clearly shows the advantage of a mobile measurement device that does not rely on fixed sites for the determination of its calibration parameters". I would assume that the main reason for the discrepancies observed with SCH measurements is due to the difference of averaging time : 9.6s response time for the QCLAS and 1min for the CLD. SCH being impacted by heavier traffic than STA , fast events of strong pollution will have more impact."

R The wording was not clear enough and has been improved.

M *"This can be expected for two reasons: (i) the larger distance, (ii) the location of site SCH at a crossroad where traffic flow is controlled by traffic lights and where the short term variability of NO_2 is larger. From a more general perspective, this is of importance when mobile sensor data is corrected based on data from nearby fixed reference sites, as suggested by (Arfire et al., 2015; Saukh et al.,* *2015). The parameters for data correction as determined from such an approach depend on the local concentration variability and the time response of the sensor and the reference instrument."*

C "4) Figure 8 : What is the meaning of points after the 7th stop ?"

R The route section shown in Figure 8 is not congruent with the tram stops, i.e. the route section is longer than the distance given by the tram stops and its length is not an exact multiple of 50 m. We suppress plotting of any boxplots for which the corresponding number of measurements is below 20. This was the case for the boxplot on the right side of the Figure to which the reviewer referred.

M The figure has been updated accordingly.

C "5) In this section, temporal variation of NO2 should be studied as a function of other parameters : NO, O3, H2O,T,solar activity (see general comments)."

R We thank the reviewer for the remark, however, the suggested study is beyond the scope of our manuscript. Here we mainly focus on the development, applicability and validation of a compact and mobile laser spectrometer for ambient NO_2 measurements. As mentioned earlier, this work does not investigate the processes leading to the formation or transformation of NO_2 and influencing the spatial variability, but rather provide a concept for spatial mapping of urban NO_2.

The statistical models directly rely on measurements that refer to confined 30 minutes time periods. The NO_2 concentration at any location is assumed to be constant within 30 minutes. Temporal variations in concentration due to changes in emissions and meteorology are captured by variations in the measurements of consecutive 30 minute time periods. The statistical models are not intended to offer a full atmospheric chemistry and transport analysis but to map the instantaneous NO_2 concentration field.

C "**i)** Model validation by means of three seconds NO 2 measurements from the tram

Should be performed with 9.6s averaging measurements"

R See our reply above.

C "**j)** 3.3.2 Model validation by means of 30 minutes mean NO2 concentrations at fixed AQM stations

1) Fig 10: change the color code of NO2 concentration to allow to distinguish between very low concentrations (presently in white) from the zone where the model cannot compute concentration. Presently, the spatial coverage is not clear."

R The figure was replaced. Areas for which no NO_2 concentrations are predicted based on the statistical models are now depicted in white and denoted as "NA". The concentrations of the model predictions are well above zero and therefore depicted in yellow (and red).

C "2) P17, L29-30 : " The obtained RMSE values range between 1.5 ppb at HEU (elevated background site) and 7.5 ppb at RGS (roadside)." Given values are not consistent with Fig 11."

R These numbers do not refer to Figure 11 but to an additional analysis that quantifies the magnitude of the accuracy of the NO_2 concentration of a location type estimated by statistical models based on measurements from one tram instrument.

In this analysis we quantified the error that may result when the 25 minute mean concentration is estimated based on the complementary measurements (5 minutes) of a 30 minutes interval (6 alterations). This error is 1.5 ppb for HEU and 7.5 ppb for RGS.

The elevated AQM site HEU is not directly impacted by traffic and encountered an annual mean NO_2 concentration of 18 ug/m^3 in 2015. Accordingly, temporal NO_2 concentration variations are small. In contrast, AQM site RGS is heavily impacted by traffic resulting in an annual mean NO_2 concentration of 504 ug/m^3 in 2015. Temporal variations in NO_2 concentration are much higher.

We rephrased this paragraph in order to better present our findings.

M *"In our statistical models, the mean of all the measurements in R_i (covering approximately 1/m of 25 minutes with mostly m = 6) is taken as the estimate value for the 30 minute mean concentration of location type R_i. We performed the following computation in order to analyze the agreement between the mean derived from a limited number of measurements and the true 30 minute mean concentration for several locations. For this analysis, measurements from single AQM sites from the year 2015 were used. We computed 5 minute means and complementary 25 minute means (6 pairs in 30 minutes, respectively). The obtained RMSE values range between 1.5 ppb at HEU (elevated background site) and 7.5 ppb at RGS (roadside). Therefore, differences between the mean of a subset of tram measurements referring to a particular location type and its 30 minute mean have to be expected if NO_2 concentration at this location is highly variable on the short-term."*

C "3) What is the error/shift value that can be attributed to the absolute calibration of the different CLDs ? Were they regularly calibrated on the same gas sample ?"

R The total uncertainty of 1.25 ppb at 30 ppb concentration was previously discussed. The CLDs of the AQM sites are regularly (25 h intervals) checked and calibrated.

C "**k)** Conclusion

1)P20L2 : " QCLAS spectrometer that operated autonomously.... over a period of four months" should be added a mention to the intervention for optical alignment

R Changed as suggested.

C "2) Please indicate clearly that the 2 ppb agreement is obtained after a correction factor of 1.2 derived from these measurements "

R Done.

C “3) 3s to be replaced by 9.6s.”

R For the reasons stated above we are convinced that using 3 second averages is appropriate.

C “Typesetting:

- Section 2.1 is missing
- please enlarge the text in the graphs or improve the resolution to allow easy reading in figure 5,6 ,7,9,10 and 11

R The original figures that were copied into the MS-Word template are in PNG or PDF-Format. We expect less compression and better quality when the figures are integrated during final typesetting.

- P1 L10,L22 and everywhere: "measures" should be replaced by "measurements" ?
- P4L13 : a word is missing "measurements are available THEY are depicted..."
- P16L9-10 : " The frequency a particular variable is effectively used" ”

R We implemented the necessary changes.

Reviewer #3

General comments:

C Further effort should be made on the quality of the writing, the English and the way of presenting data. The paper structure should be reviewed: The main structure: Introduction, methods, results and discussion and conclusions is ok, but the sub-structure should be reviewed. The section 2.4 should be partially included in section 2.3 and 2.5. Section 3.1.should not be split into two sections, same for section 3.3. Finally in figure 10 authors could add the name of the stations so that readers don’t need to go to the supplementary material to see station’s positions. Finally, authors are making data analysis with 3 seconds averaged data while the instrument response time is larger (9.6 seconds reported in page 3 or 11 seconds in page 9). The whole data analysis should be redone accordingly.

R We appreciate the general comments. We are convinced that using sections and subsections increases the readability and therefore prefer to keep the structure of the manuscript as is.

Figure 10 is updated also following the comment of Reviewer #2.

The flushing time of the instrument and the temporal resolution used in the data analysis are addressed in detail in response to reviewer #2’s comments. In short: Sampling at a higher frequency than the flushing time is necessary. But we were mistaken to add the flushing time as a delay to the instrument response and therefore redid the data-analysis and statistical modelling without that delay.

Specific comments

C Page 2, Line 8 : « With CLDs, NO2 is determined indirectly as the difference of consecutively measured NOx and NO, making this technique not fast enough for mobile measurements ».

To my knowledge CLDs can switch between NO and NOx channel within only few seconds, which I believe is still considered high resolution (<1 hour as mentioned at page 1 line 28). What it can happens is that in case of rapid NOx transitions the value of NO2 can be biased, leading to overestimation or underestimation of NO2 content. Furthermore, since is an indirect approach, it can suffer from artefacts (for instance during the conversion NO2 -> NO in the catalytic oven, other species may also be converted, leading to an overestimation of NO2). I think authors should reformulate this part more in those lines, and please add references.

R The reviewer is correct that CLDs can switch between NO and NOx channels within a few seconds, however, instruments typically use a low pass data filtering algorithm leading to a much slower time response. For example, the response time of the CLD used within the Swiss Air Pollution Monitoring Network (Horiba APNA-360) is 120 seconds which is not fast enough for mobile measurements (for example on a tram). “High resolution” as mentioned in page 1 line 28 refers to the temporal resolution of a pollution map.

We agree with the reviewer concerning the sources of uncertainty of CLD NOx measurements and will modify the section accordingly.

M *“The indirect measurement approach of the CLD instruments can lead to biased NO_2 measurements when NO and NO_2 are temporally highly variable. Furthermore, the conversion of NO_2 to NO in a catalytic converter can lead to measurement errors as other species such as NH_3, HNO_3 and HNO_2 can be converted, too, and therefore lead to NO_2 signals that are biased high (NABEL, 2016; Steinbacher et al. 2007).”*

C Page 3, Line 4: « Light passes a 12m … the infrared detector ». Move this below together with the cell description. Could authors comment on the possibility to reduce the cell volume? If my calculation are correct, the height of the cell is 1.8 cm, while usually the beam waist for such a type of propagation is 4-5 mm. It could be possible to reduce the height in order to gain a factor of 2 on the cell volume or there are other factors that will make this volume ideal?

R We moved the cited sentence as suggested. In principle, the reviewer is right: If all reflections were in the same plane the cell volume could be reduced by half the volume. However, the off-axis configuration was found to be favorable in terms of optical noise (fringes). As such, the reflections alternate between two horizontal levels on the mirror, and thus, there is very little room for further reducing of the cell volume.

C Page 3, Line 17: « We determined the flushing time for a complete gas exchange in the cell to be 9.6 s by exponentially fitting the NO2 signal after switching from outside air to filtered NO2 free air». Is that what expected according to the total volume of the instrument and the gas flow at 100hPa of pressure? Do you see any memory effects that slow down the gas exchange process? Do you observe the same response time while going from outside air to filtered air and on the other-way around?

R We detailed the gas flow and exchange time in a response to *comment c)4)* of reviewer #2.

Yes, the response time is the same in both directions.

C Page 3, Line 29: « the instrument repeatedly determines the zero-point offset measuring filtered NO2 -free air ». Mention already here how often this zero-point is determined.

R Done.

M *"…by measuring filtered* NO_2*-free air for 2 minutes every 20 minutes."*

C Figure 1a: It could be more interesting to have a picture of the instrument without the cover to be able to see the inside; Figure 1b: Further efforts should be made to provide a clearer and more informative schematic of the setup.

R We agree and exchanged the figures also in accordance with comments of reviewer #2.

C Page 4, Line 13: « The gap between 27 Dec 2015 and 05 Feb 2016 is due to discarded data of reduced quality related to misalignment in the optical system of the instrument.». It could be interesting to know how many time the instrument required the intervention of the operator over this period of time to show the robustness of the instrument. Any explication on why it took so long (more than 1 month) to realign the optical system?

R From 27 Dec 2015 to 05 Feb 2016 there was only one intervention by the operator: The instrument was removed from the tram and taken to the laboratory on 27th January 2016 and installed again on the tram on 4th February 2016 after realignment and testing. It did not take one month to realign the instrument, but the data from 27th December to after the repair had to be discarded.

C Page 4, Line 15: « Moreover, data was omitted when the NO2 spectrum was not clearly identified by the processing algorithm ». Authors should specify the percentage of "bad" data that were omitted. Is that related to S/N ratio? Did authors used a threshold on the standard deviation of the fit? More explanations are needed for this point. Replace "data was omitted" with "data were omitted".

R In total 37 % of the measurement days yielded data that could be used for statistical modeling. We used the water concentration derived from the measured spectrum to flag measurement periods where the fitting algorithm did not work correctly. The according information is added in the manuscript.

M *"The water concentration determined from the measured spectrum was used as an indicator to flag such data. In total 37 % of the measurement days yielded data that could be used for statistical modeling."*

C Figure 2: « high quality NO2 data ». Authors should be more clear in the manuscript about what good quality data means. Which rules have been used to reject data?

R The rules for rejection of data are stated in section 2.2.

Data were used if…

- …the tram is not in the depot
- …its position can be attributed to the track
- …the NO_2 spectrum is clearly identified (which is the case when the water concentration measured alongside the NO_2 is between 0.1 and 4%)

C Section 2.3: From making the reading more clear, please specify the different stations (ZUE, SCH, HUE etc) and mention which one are in a background zone and which one in a polluted area.

R In response to *comment e)* of Reviewer #2 we added information about the annual mean NO_2 concentrations at the different sites to section 2.3

M *The reported annual mean NO_2 concentrations in 2015 for the AQM sites were: HEU: 18 µg/m^3, ZUE: 31 µg/m^3, STA: 33 µg/m^3, SCH: 45 µg/m^3, SWD: 47 µg/m^3, RGS: 50 µg/m^3.*

C Page 5, Line 18: « Largest emission sources contributing to ambient NO2 concentrations in Zurich are motorized traffic (47% of NOx) and heating systems (28% of NOx) ». What is the remaining 25%?

R The missing 25% are: 15% industry, 9% constructions and 1% agriculture and forestry ("Luftbilanz der Stadt Zürich", 2012). We added the information to the manuscript.

M *Largest emission sources contributing to ambient NO_2 concentrations in Zurich are motorized traffic (47% of NO_x) and heating systems (28% of NO_x) (followed by industry (15%), constructions (9%) and others (1%)) (Brunner and Scheller, 2014). There is no heavy industry.*

C Page 6, Line 8: "We used four types:" of what? The object is missing.

R The sentence should be: "We used four types of traffic intensities:"

M *…of traffic intensities:*

C Page 8, Line 11 : « Such a correction to the normally calibration free QCLAS method ». Authors should also discuss and if relevant quantifying the effect of missing UV radiation in the inlet of the instrument that would perturb the photo stationary state ratio between NO and NO2. They mention a flushing time of 11sec that is long enough for accounting for an overestimation of the NO2 (due to the reaction of NO with O3 in absence of UV light). A quantification of this artefact should be reported.

R As mentioned in our response to reviewer #2, sampling artefacts, e.g. through formation of NO_2 in the inlet from reaction of NO with O_3 is inherent for every measurement using a closed measurement cell. The residence time of the air in the sampling line and the flushing time in the measurement cell have been kept as short as possible, so that the sampling artefacts are minimized. The residence time for the reference CLD instruments is comparable (about 6s in sampling line plus mixing time in measurement cell), so it can be expected that the sampling artifacts of the reference NO_2 measurements (which are performed in agreement with the European Norm EN 14211 which is mandatory for regulatory measurements) are comparable to the QCLAS measurements.

A quantification of this measurement artefact is not possible because the mobile measurement platform was not equipped with NO and O_3 measurements. However, a back of the envelop calculation suggests that the artefact is below 1 ppb (assuming the rate constant for the reaction of NO and O_3 at 298 K being 0.0004 1/(ppb·s) and NO and O_3 concentrations of 20 ppb and 10 ppb, respectively – note that the measurements have been done during the cold season at low O_3 concentrations).

M We added the following sentences: *"Note that formation of NO_2 in the sampling line of instruments due to reaction of NO with O_3 can lead to an overestimation of NO_2. The influence of this artefact is small, especially because the measurements have been done during the cold season at low O3 concentrations. It is, however, for the QCLAS a bit higher than for the reference CLD instrument, because the residence time of the air sample is for the CLD somewhat shorter (about 6s)."*

C Page 8, Line 17: "Discrepancies between the two instruments are mostly due to the higher temporal resolution of the QCLAS and imperfect time synchronization." This should be avoided if QCLAS data are averaged according to the acquisition rate of the CLD (which seems to be the case as mentioned in the caption of figure 4 (1 minute average for both dataset). Were time stamps of the two instruments synchronized before starting the comparison? This statement should not be necessaire if data comparison is done properly.

R We agree that the above statement might be a bit confusing. We are comparing 1 minute average values of a CLD and the laser spectrometer. The CLD has a response time of 60 seconds and the laser spectrometer has 9.6 seconds flushing time. Single pollution events of less than 60 seconds duration cannot be captured in the same way and therefore lead to different concentration signals. The time synchronization was done properly and therefore imperfect time synchronization cannot explain discrepancies.

M Text has been changed accordingly.

C Page 9, Line 1: "The standard deviation of the QCLAS measurements from the CLD was found to be 0.96 ppb therefore we conclude that the instrument accuracy is about 1 ppb and that the assumption of a constant relative loss of NO2 was correct." Authors should mention the precision of the CLD measurements. This sentence is not clear: is the 0.96 ppb the standard deviation of QCLAS data with respect to CLD data as plotted in figure 4. If so, the sentence "The standard deviation of the QCLAS measurements from the CLD" should be written more as: "The standard deviation of the QCLAS measurements with respect to CLD data". Further discussions about the sensibility of the instrument should arise here. According to the Allan-Werle analysis (Fig.3) the precision of the instrument should be0.02 ppb. Since that is not the case, this means that the AW-deviation should start to go up in a longer term. I believe that authors should provide an AWW-deviation plot on a longer term to report this long term drifts that will justify this final precision of 1 ppb.

R In response to *General comment 5)* of Reviewer #2 we stated the precision of the CLD that was used for the comparison and we will rephrase as suggested.

The Allan-Werle plot (Fig. 3) was updated with a larger data set to show the drift of the instrument after about 400 seconds.

1 ppb is the accuracy of the laser spectrometer in comparison with the CLD of the AQM site after calibration and zero point correction.

M Done as suggested.

C Page 9, Line 10: "We discarded all the positions with corrections exceeding 10 m". Authors should mention the percentage of rejected date due to inconsistent position. Was 10 m decided because of the optimal distance for high resolution data? Was this threshold fixed also according to the GPS position? Discussion on this choice will be appreciated.

R The accuracy of GPS measurements in densely built environments can be deteriorated due to the reception of signals from fewer satellites than maximally possible and due to multipath effects in GPS signal propagation. The 10 m threshold is arbitrary but this approach is straightforward. GPS position accuracy can be expected < 5 m under good conditions. The tram is bound to the tracks. The larger the distance between GPS position and the nearest tram track the less reliable is the GPS position. The reliable assignment of the NO_2 measurement to a position is required so that the data can be used in the statistical models.

M We suggest to not to extend the corresponding section.

C Page 9, Line 13: "we added 11.1 seconds to the timestamps of the matched positions in order to account for the flushing time of the measuring cell and for half of the length of the measurement interval." What it should be taken into account is renewing of the sample in the cell, which corresponds to the residence time of the gas in the inlet plus the residence time in the cell. It is not clear here if the 11.1 sec correspond to this. Why half of the length of the measurement interval is also taken into account? At page 3 the response time was estimated to 9.6 sec. Please change the one that is not right.

R In the data analysis of the initially submitted version of the manuscript we subtracted a time constant of 11.1 seconds from the time of the NO_2 measurement and related the NO_2 measurement to the position at this time. The time constant was the sum of 1.5 seconds (the half of the 3 second averaging period) and 9.6 seconds (time constant for NO_2 changes in the measurement cell). We changed the data analysis in the revised manuscript by omitting the 9.6 seconds.

M *"Hereby, we added 1.5 seconds to the timestamps of the matched positions. Such a delay is necessary because at a given time-stamp the GPS provided the present position while the* NO_2 *concentration is the average of 3 seconds before a time-stamp."*

C Page 9, Line 15: "The NO2 measurements refer to 3 seconds averaging time. Therefore, they are strictly speaking not point measurements but refer to a route segment." Not well written. Please rewrite it with a better English. And as mentioned above 3 seconds average is not ideal because the instrument response time is larger.

R Done.

M *"The* NO_2 *measurements represent 3s averaged data, which, considering the traveling speed of the tram, cover a given route segment instead of representing a fixed point measurement."*

C Page 9, Line 17: "The length of this segment is the product of the integration time of 3 seconds and the speed of the tram. Tram speed associated with the mobile NO2 measurements was smaller than 5 m/s for 64 % and smaller than 10 m/s for 92 % of the time." What is important here is also the average speed of the tram. This will also question the 10 m threshold that authors selected for discharging data because of bad positioning. If the measurement is done in 3 sec and the maximum tram speed is above 10 m/s then the a threshold of 10 m on the position data is not justified and it should be higher in my opinion (It should corresponds to the maximum speed of the tram times the measurement time. eg. 10m/s x 3 s = 30 m threshold).

R The distance between tram stations in Zurich is normally a few 100 meters. The actual tram speed depends on track characteristics (e.g. curves, traffic lights) and maximum acceleration rates. At each station the tram stops for the entry/exit of passengers. Additional stops may occur at traffic lights.

The GPS receiver on the tram operates independently from the QCLAS instrument. First, outliers in the GPS positions have to be identified. If a GPS position is located more than a distance D away from the tram track the true position error is equal or larger than D. 10 m corresponds to about three times the accuracy of a single GPS position. Second, the GPS positions were orthogonally projected on the tram tracks as the tram is bound to the tracks. Third, the position of the NO_2 measurement is interpolated from the improved GPS positions.

This means:

1. Continuous measurement of NO_2 with QCLAS in (ti,ti+3s) → Average NO_2 concentration for (t_i+1.5s)
2. Improved GPS position at t_k, t_k+3s, t_k+6s, …
3. Interpolation of tram position at t_i+1.5s

Obviously, the error of a single GPS position impacts the resulting position of the NO_2 measurement and also the link between the NO_2 measurement and the spatial information used in the statistical models.

M We suggest keeping the text as it is.

C Page 10, Line 3: "We analyzed the quality of the data obtained in the mobile application twofold:

First, we quantified the measurement uncertainty related to noise and variations of the zero point offset. Second, we compared the tram measurements to measurements from a fixed air quality monitoring site (see next section)." What is the mobile application twofold? The whole sentence is not clear and should be rewritten.

R Twofold means: "In two ways" or "twice". We modified the text accordingly to avoid confusion.

M *"We analyzed the quality of the data obtained in the mobile application in two ways."*

C Page 10, Line 14: "We corrected the NO2 measurements by linearly interpolating the zero point offsets derived in two consecutive zeroing periods… … … Zero point offsets are temporally correlated as the standard deviation of the differences of two consecutive zero point offsets amounts to 2.1 ppb and thus is smaller than 4.3 ppb expected for white noise." I think that this part should be better written. I understand the idea of "predicting" the zero point offset using the two neighbour ones. What is not clear is: is this analysis used to remove "bad" zeroing point? If so, it should be expressed more clearly. Where the 2.1 ppb comes from? From my understanding zero point offsets will be temporally correlated if the difference between the predicted and the measured zero point value (which I will not call standard deviation) will fit within the 4.3 ppb expected in case of white noise. The quantiles are highly reported in this work but not enough used in the discussion.

R We computed:

- Zero point offset at time i: ZPO (ti)
- Difference of two consecutive ZPO: dZPO

If the ZPOs were normally distributed, their differences would be normally distributed as well and the standard deviation of the differences would be sqrt(2)*sd(ZPO)=4.3 ppb. However, we found a standard deviation of only 2.1 ppb for the differences showing that consecutive ZPOs are temporally correlated. Temporal correlation is a requirement for the linear interpolation between the ZPOs.

The quantiles are reported in addition to the histograms in order to support the reader in comprehending the distributions shown in the figures.

M We suggest to keep the text as is.

C Page 11, Line 8: Why authors compare measurement from the tram and fix site with different time averages? For fig 6 is clear that measurements with 3 sec integration can show much high variability (due probably to local production). To be more consistent, the comparison should be for the same time window and same averaging time. Authors should discuss about the higher variability observed at shorter time averaging (3 sec).

R In fact, figure 6 shows 5min averaged NO_2 concentrations of the sites STA (red line), SCH (blue line) and 5min averaged concentrations measured by the tram instrument (black line). In this way the data can be well compared. Additionally, the data of the tram instrument is shown in full temporal resolution (grey). But the message of this plot is that the mobile instrument that regularly passes both fixed sites is in good agreement with the fixed sites when the same averaging is applied. We suggest keeping the text as it is.

C Page 11, Line 14: The statement "The calibration of sensors using precise measurements from AQM sites is an option for mobile operated sensors" requires better explanation.

R We added explanation to the respective section in response to *comment h)3)* of Reviewer #2 and now it should be clear.

M See reply to comment of Reviewer # 2.

C Table 1 Caption: replace number of values with number of measurements. Here too, the quantile values are not well used in the discussion. Authors should either used the more critically or remove them from the table**.**

R Thanks, caption has been changed. We suggest keeping the quantiles in the table, as they provide insights in the variation of the differences of mobile tram based measurements and the fixed air quality monitoring sites.

M Caption has been changed.

C Page 11, Line 21: Suggestion to replace "aggregated by 50 meters of track length" with "with 50 meters resolution". Same in the Caption of Figure 8. As mentioned in the caption, the fact that in the figure are the results of the full campaign (and not only 1-way from Bellevue and Bahnofquai should be also mention in the text. Authors should give an idea if how many 1-way trips are used for this plot and how much this corresponds in term of time.

R We suggest to keep only the key messages in the text and leave the more technical information about the number of underlying measurements (148 runs) in the figure caption.

M Text has been changed as suggested.

C Figure 8. "The tram is directly impacted by traffic between the first and the third stop as well as between the fifth and the seventh stop. There is only little traffic between the third and the fifth stop." Not important in the caption, but this discussion should come out in the text, mentioning that there is then a correlation between the zones which are highly impacted by traffic and the

NO2 level measured. What are the "extra" dots in the graph? Out-layer data? This should be specify.

R We agree. Text and figure caption will be changed as suggested.

M Text and caption of Figure 8 has been changed.

C Page 11, Line 21: what authors mean for similar locations? This is something that it should be explained more clearly. What are the criteria for making two locations similar?

R We reworded the sentence to improve understanding.

M The text has been changed to: *"Our statistical models are driven by measurements and are based on the assumption of comparable NO_2 concentrations at locations that are similar in terms of vicinity to sources (in particular traffic), the built environment and elevation."*

C Figure 9. There is a miss of general explanation about this analysis. I have the feeling that all for plots at figure 9 don't bring a lot of information from the statistical point of view. I think authors should do a stronger effort in reporting data which are important for the understanding of the work. In the text this figure is not well descripted, ant the readers don't really get the point of what is plotted. Few more details: The really last red dot at the top right is suspicious… For this measurement period there should be a black dot (MAXTi) on top of the Q095Ti dot. Modelling period is a bit confusing because to me this is a simple data analysis, and there is not modelling behind. I suggest to change it with measurement period. The use of "Frequency" on the y-axes to me is also ambiguous, I would suggest to put "number of occurrences". I found weird that in figure c and d the y-axes does not extend to include the whole data set.

R In this case, we do not agree with Reviewer #3 and think that Figure 9 provides useful information about the spatio-temporal variability of NO_2 measurements in the city of Zurich, as obtained from the mobile measurements. It is true, Figure 9 is somewhat dense and complicated but we feel that it is appropriately presented and explained in the text. We therefore suggest to leave Figure 9 and the corresponding discussion in the text as it is.

M No changes.

C Page 16, Line 9: "The frequency a particular variable is effectively used for splitting the 30 minutes data set depends only on the measurements". I guess authors wanted to say that "The frequency of a particular variable is effectively used for splitting the 30 minutes data set depending only on the measurements". Anyway this sentence should be rewritten more clearly.

R The reviewer is right. We could rephrase it as suggested but think that the sentence is not needed at all. It does not provide additional information for the reader.

M Sentence has been deleted without replacement.

C Page 16, Line 11: The whole paragraph: "The number and spatial distribution of measurements in a modelling period (approximately 500)… …Regression trees based on tram measurements have about 6 location types Ri on average." should entirely be rewritten. I don't see how the number of measurements in the modelling period will limit the possibility of separate different location types?

R The concentration predicted by the statistical models is the mean of all the measurements referring to this location type. The higher the variability of NO2 concentration at a particular location the more measurements are required to accurately estimate the mean value. A representative set of measurements has to be assigned to each location type.

M *"The number of location types for which representative NO_2 concentrations can be determined based on our statistical models is limited. The number of measurements required for estimating an accurate mean concentration for a location type depends on the short-term temporal variability of the NO2 concentration (see Figures 9 (c) and (d) and section 3.3.2). However, the number and spatial distribution of the measurements in a modelling period (approximately 500) is given by the tram route and cannot be influenced. Accordingly, the use of too many explanatory variables in our statistical models may lead to an over-determined model. Regression trees based on tram measurements have about 6 location types Ri on average."*

C Section 3.3.1: I am sceptic on meaning of the model version Vt, especially on a such a short time scale. I guess that this approach could be les redundant by increasing the time intervals. For instance using the 30 min intervals to predict other 30 min intervals. Add "respectively" at the end of sentence (line 23).

R The statistical models rely on measurements and do not include any modeling features based on dispersion or chemistry. We selected the modeling time period to be 30 minutes in order to minimize effects from changing emission sources and meteorology. As models Vt perfom best in this modeling approach we concluded that there might indeed be potential for denser air quality measurement networks. We think that the conclusions are adequate to the performed analysis.

We think that the comparison of the model is correct and adequately described and suggest to keep the text as it is.

M Line 23: The word "respectively" has been added as suggested.

C Page 16, Line 28: "The difference in performance between VAQ and VSTA is small due to the similarity of NO2 concentration encountered at site STA and by the tram-based instrument." Why? VAQ and VSTA are both based on fixed site. How come their similarity should be due to the similarity between NO2 measurement at STA and on the tram?

R While VSTA is based on data of AQM site STA only, VAQ uses data from all AQM sites including data from AQM site STA. Environment along tram tracks is most similar to that of STA in terms of traffic impact. That is why NO_2 concentrations of site STA are taken in the modelling as predictions for the NO_2 concentration for most tram positions. This explains why versions VAQ and VSTA only slightly differ.

C Table2: How the variables to be used on each model have been chosen? Why there are not variables for VSTA? Same question for section 3.3.2. From what authors describe it seems that variables are chosen to achieve the best overall fit to the measurements at AQM sites. But those variables have a specific meaning and they should not be selected to making the measurement better matching but more according to their physical meaning.

R VSTA is based on data from site STA only (one site or location). Partitioning of the data in terms of geographic information is not possible.

We think that the variables used in the models (traffic intensities, SVF and elevation) have a relation to the instantaneous NO_2 concentration field in Zurich. For example, the dependence of NO_2 concentration on distance to a road depends on meteorology (beside other factors such as the built environment). Therefore, traffic intensities with different decay constants are included in the models. The idea of statistical modelling is to directly relate measurements with information that characterize locations. By validation methods such as leave-one-out the accuracy and limits of the derived models are tested.

C Figure 11 and its discussion: In my opinion, Figure 11 shows that model prediction is less effective at stations far away from the tram track (BLU and HEU). As expected, predictions that use data from the stations (WAQ and WT+A) provide better RMSE but still slope values are more far from unit. I don't think difference between models can be really discussed here. What about error bars in figure 11?

R The reviewer is right that the predictions based the statistical models are less accurate if a location largely differs from the locations of the measurements the model is based on in terms of traffic impact, building density or elevation. That is true for models based on mobile measurements as well as for models based on fixed AQM sites.

We compare in Figure 11 models based on the tram measurements, measurements from AQM sites as well as their combination. The fact that the model versions VAQ and VT provide similar results is important as VT is based on only one mobile instrument. Figure 11 further shows that if the spatial NO_2 concentration is derived by a combination of measurements and statistical modeling the set of measurements must provide a good coverage of all the pollution situations. The AQM sites operated in the city of Zurich provide a good picture of the pollutant situation in Zurich although the differentiation of pollutant types is limited by the number of stations.

M We suggest to keep the text as it is.

C Section 3.4: "Our results suggest that the number of required mobile or static QCLAS instruments for comprehensive NO2 measurement in Zurich is not smaller than the current number of fixed AQM sites." How authors get to this conclusion?

R Section was probably unclear. We tried to make this point clearer.

M Following text has been added: *"Due to the number of required measurements for location types with highly variable* NO_2 *concentration, the necessity for measurements at locations with different pollutant situations and the travel time from one to another location, we suggest that the number of required mobile or static QCLAS instruments for comprehensive* NO_2 *measurement in Zurich is not smaller than the current number of fixed AQM sites."*

C Section 3.4: What about taxis? I guess those could provide a more random and spread-out data set with respect to public transportation which will follow a fix path.

R Yes, passenger cars could also be equipped with QCLAS sensors. However, a Tram or Bus has the advantage that an instrument can be installed permanently and sufficient power is available.

M No changes.

C Page 19, Line 19: "The model type the maps are based on is irrelevant in this context." Meaningless.

R We want to emphasize that the use of such an instrument for validation is independent from statistical models. However, we realized that this sentence can be misleading and is not important here. We therefore suggest to delete this sentence without replacement.

M Sentence has been deleted without replacement.

C Page 19, Line 23: "within the same footprint." What does it mean? Same source?

R We meant that with such lasers we could build an instrument of the same size that measures NO and NO_2 simultaneously. We will change the wording.

M Text has been changed to "*With the latest dual-wavelength QCL technology (Jagerska et al., 2014; Süess et al., 2016), the next generation of the QCLAS instrument could measure NO and* NO_2*, simultaneously, with the same compact geometry.*"

C Technical comments:

Page 2, Line 10: « Similarly, electrochemical sensors are currently not suited for such applications as their response time is not short enough ». Add a reference

Page 2, Line 25: « a cylindrical multipass cell ». Put the reference Mangold 2016 also here.

Page 2, Line 27: replace 1.1 with 2.1 Page 2, Line 30: remove Additionally

Page 3, Line 14: change "it is not sensitive" with "it is less sensitive".

Page 3, Line 17: « while the air flux is limited by a 100 µm diameter orifice ». Specify to which flow.

Page 3, Line 22: « nano PC ». Specify which one.

Page 3, Line 34: Say what GSM stays for.

Page 4, Line 8: « altering intervals ». I suggest to replace altering intervals with varying intervals.

Page 6, Line 5: replace "in the environment of this location" with "in a specific location".

Page 6, Line 17: replace « for every three seconds » with “every 3 seconds ”.

R We thank the reviewer for the technical comments and made the necessary changes.

Spectroscopic real-time monitoring of NO_2 for city scale modelling

P. Morten Hundt[1], Michael Müller[1], Markus Mangold[1,3], Béla Tuzson[1], Philipp Scheidegger[1], Herbert Looser[1,2], Christoph Hüglin[1] and Lukas Emmenegger[1]

[1]Laboratory for Air Pollution and Environmental Technology, Empa, 8600 Dübendorf, Switzerland
[2]Institute for Aerosol and Sensor Technology, FHNW, 5210 Windisch, Switzerland
[3]*Present address*: IRsweep AG, 8093 Zürich, Switzerland

Correspondence to: P. Morten Hundt (Morten.Hundt@empa.ch)

Abstract. Detailed knowledge about the urban NO_2 concentration field is a key element for obtaining accurate, individual exposure estimates. These are required for improving the understanding of the impact of ambient NO_2 on human health and for related air quality measures. We developed a compact and robust quantum cascade laser absorption spectrometer (QCLAS) and deployed it on a tram in the city of Zurich (Switzerland) to perform mobile real-time concentration measurements of NO_2~~.~~ by direct absorption spectroscopy at 1600 cm^{-1}. Thorough analysis of the obtained NO_2 data, for instance by comparison with data from fixed air quality monitoring (AQM) sites, revealed the instrument to be highly accurate and valuable for collection of data that can be used in statistical models for the calculation of spatio-temporally resolved NO_2 concentration maps. The combination of fast mobile measurements with AQM data proved to be very suitable, but the statistical data analysis also showed that a single mobile instrument is not sufficient in the studied urban area, for mainly two reasons: (i) short residence close to sources with large short-term NO_2 variations and (ii) limited representativeness of the tram tracks for the entire urban environment.

1 Introduction

Numerous studies relate the exposure to nitrogen dioxide (NO_2) to adverse health effects (e.g. (Adam et al., 2014; Gehring et al., 2013; WHO, 2013)). Despite this threat to human health, limit values are regularly exceeded in European cities, mainly at locations directly impacted by traffic emissions (EEA, 2016). It is, therefore, highly relevant to provide spatially and temporally resolved NO_2 fields for the assessment of related health effects, for the guidance of efficient air quality measures and for urban air quality planning. True exposure of an individual is composed of the encountered pollutant concentration at a particular location and time and breathing rate. So far, the individual's exposure is mostly derived from Land Use Regression models representing seasonal or annual mean concentrations and the individual's home (and working) address (e.g. (Brauer et al., 2008; Cyrys et al., 2012)). Obviously, these values may be significantly biased depending on the mobility pattern of an individual as NO_2 concentrations in the urban environment are highly variable in space and time.

Spatially and temporally highly resolved pollution maps (< 20 m, < 1 hour) based on statistical modelling can enhance the accuracy of exposure estimates. However, such statistical models require accurate input data that represent the concentration levels at a set of different locations in an adequate temporal resolution. Until now, accurate (±1 ppb) and continuous NO_2 measurements are mainly performed at air quality monitoring (AQM) stations equipped with chemiluminescence detectors (CLD). Data from such AQM stations (e.g. seven locations in Zurich) do not provide sufficient spatial resolution for detailed pollution maps due to the high variability of influencing factors such as the traffic situation and the built environment.

One approach to overcome this challenge is the deployment of denser measurement networks. Mueller et al. (Mueller et al., 2015), for instance, used data from a dense network of passive diffusion samplers in Zurich (Switzerland) as input for statistical models. Such passive samplers, however, provide only average values over their exposure periods, which are days to weeks, and therefore lack the necessary temporal resolution. The replacement of passive samplers by low-cost electrochemical sensors is being investigated but has not yet shown to be feasible for long-term deployment (Mead et al., 2013).

Mobile measurements are another option for increasing the spatial resolution (Hagemann et al., 2014; Hasenfratz et al., 2015; Kehl, 2007). They require ~~fast and~~ highly sensitive – typical concentrations range from 5 to 200 pbb – and selective measurement devices with temporal resolutions of just a few seconds. With CLDs, NO_2 is determined indirectly as the difference of consecutively measured NOx and NO, making this technique not fast enough for mobile measurements~~.~~ ~~Similarly, electrochemical sensors are currently not suited for such applications as their response time is not short enough. Mid-infrared laser spectroscopy, in contrast to CLD, determines the concentration of NO_2 directly by measuring the absorption by ro-vibrational transitions of the NO_2 molecule. Therefore, the measurement is highly selective and because of the relatively high absorption cross sections in the mid-infrared, it is also very sensitive with detection limits in the range of a few ppt.~~ as their response time is longer than 1 minute. The indirect measurement approach of the CLD instruments can lead to biased NO_2 measurements when NO and NO_2 are temporally highly variable. Furthermore, the conversion of NO_2 to NO in a catalytic converter can lead to measurement errors as other species such as NH_3, HNO_3 and HNO_2 can be converted, too, and therefore lead to NO_2 signals that are biased high (NABEL, 2016; Steinbacher et al., 2007). Also electrochemical sensors are currently not suited for such applications as their response time with about 60 seconds is too long. Laser spectroscopy, in contrast to CLD and electrochemical sensors, determines the concentration of NO_2 directly and with sampling rates of 1 Hz and higher. Previous studies have shown the applicability of laser diodes in the blue spectral region to measuring NO_2 by various cavity enhanced techniques (Courtillot et al., 2006; Fuchs et al., 2009; Kebabian et al., 2005; Osthoff et al., 2006). With higher absorption cross sections than in the visible the mid-infrared region is often chosen for high precision trace gas monitoring of NO_2. Here, for instance photoacoustic spectroscopy (Pushkarsky et al., 2006), quartz-enhanced photoacoustic spectroscopy (Patimisco et al., 2014) and direct absorption spectroscopy (McManus et al., 2015) were applied.

Our method of choice is mid-infrared direct laser absorption spectroscopy at 1600 cm^{-1} where the concentration of NO_2 is determined directly by measuring the absorption by ro-vibrational transitions of the NO_2 molecule. This method is highly selective and very sensitive with detection limits of only a few ppt (McManus et al., 2015). Tuzson et al. (Tuzson et al., 2013), for example, deployed a two laser quantum cascade laser spectrometer (QCLAS) on the high-altitude air monitoring site Jungfraujoch (3580 m a.s.l., Switzerland) to measure background concentrations of NO and NO_2. Additionally, laser spectroscopy allows high sampling rates as shown e.g. by Jagerska et al. (Jagerska et al., 2015) for simultaneous NO and NO_2 measurement in engine exhaust gas with a sampling rate of 10 Hz. Furthermore, compared to other spectroscopic techniques, direct absorption spectroscopy relies on a relatively simple optical layout comprised of a light source, a long-path cell and an infrared detector.

~~Our~~Therefore, our approach to perform accurate, mobile and direct measurements of NO_2 is to deploy a compact and robust quantum cascade laser (QCLAS) spectrometer on the roof of a tram that is operating as public transport service in the city of

Zurich. The NO_2 concentration is measured by direct absorption spectroscopy at 1600 cm^{-1}. In this article, we present our newly developed laser spectrometer, analyze the performance of the instrument and quantify the accuracy of the NO_2 measurements. Moreover, NO_2 measurements from the single mobile instrument are investigated with respect to its use in highly resolved statistical models. Our study ~~utilizes~~uses the extensive data set from air quality monitoring facilities in the city of Zurich. The information about instantaneous NO_2 concentrations obtained with fixed air quality instruments provides the opportunity to validate the tram based measurements as well as the model predictions.

2 Methods

~~1~~2.1 Quantum cascade laser spectrometer

The portable instrument was specifically built for this campaign. Figure 1 shows a photograph of the QCLAS instrument without cover and a schematic of ~~its main components~~the instrument and the optical layout. The whole instrument is built in a 40 × 36 × 15 cm waterproof box and mounted on four metal springs which absorb vibrations from the moving tram. ~~Additionally, the~~The optics are mounted on a carbon fiber breadboard which is supported by additional metal springs to further reduce vibrations. The instrument weights about 10 kg and is powered by a 12 V connection supplied by the tram. NO_2 concentrations are determined by measuring the direct absorption signal of a single ro-vibrational transition in the mid-infrared. The main components of the instrument are a single mode (distributed feedback) DFB QCL (Alpes Lasers), a cylindrical multipass cell ~~and a thermoelectrically cooled MCT detector (PVI-4TE-6, Vigo Systems).~~(Mangold et al., 2016) and a thermoelectrically cooled MCT detector (PVI-4TE-6, Vigo Systems). The laser is packaged in a high-heat-load (HHL) housing, where a Peltier element is used for temperature control. The QCL is operated at room temperature and emits 3 mW at 1600 cm^{-1}. It is driven in intermittent continuous-wave (iCW) mode (Fischer et al., 2014) with a pulse duration of 160 µs and a duty cycle of 50 %. ~~Thereby,~~In iCW mode the driving current of the QCL is dropped to zero between individual pulses leading to rapid heating and therefore frequency tuning during each pulse. The pulses are generated by discharging a capacitor over the QCL. A frequency tuning of >1 cm^{-1} was achieved (see Figure 1(c)). By shaping the current ramp with different RC-elements, close to linear tuning could be achieved. See (Fischer et al., 2014) for details. In this driving mode the heat dissipation of the laser is lowered such that a fan is sufficient to cool the laser housing, and thus no additional cooling liquid circulation is needed. ~~The light passes a 12 m optical path within the cylindrical multipass cell before it hits the infrared detector.~~

The output of the QCL is collimated inside the HHL housing by a f=1.873mm aspheric lens and focused into the multipass cell at half way of the first pass with a CaF_2 lens (f=150 mm). For frequency calibration a 5.08 cm Ge-etalon can be inserted between the laser and the first mirror. The light passes a 12 m optical path within the cylindrical multipass cell (89 reflections) before it leaves the cell through the same hole and is steered to the infrared detector. The cell is described elsewhere in detail (Mangold et al., 2016) and only a brief summary is given here. As a cylindrical cell is always concentric in tangential direction, we chose the cell to be confocal in sagittal direction which leads to a more stable beam propagation in both directions compared to a concentric arrangement. Furthermore, the mirror curvature was chosen to be parabolic in sagittal plane for better refocusing properties of the beam. To minimize interference fringes an absorption mask with holes of 4 mm diameter, where laser reflections are expected, is inserted in the cell. Additionally, the beam is coupled into the cell in an off-axis configuration leading to further separation of neighboring reflections on the mirror and thus less interference fringes. ~~The laser beam was focused into the cell at half way of the first pass with a CaF_2 lens (f=150 mm) and leaves the cell through the same hole after 89 reflections.~~ With a diameter of 14.5 cm the cell has a volume of 300 ml. In contrast to alternative multipass-cell concepts, like Herriott-cells, the cell is built from a solid ring. Therefore, it is ~~not~~less sensitive to internal misalignment due to shock or vibrations.

In order to obtain narrow, well separated absorption lines, the pressure in the cell is reduced to 100 hPa by continuously pumping on it while the air flux is limited to 180 sccm by a 100 µm diameter orifice placed at the gas inlet. We determined the flushing time for a complete gas exchange in the cell to be 9.6 s by exponentially fitting the NO_2 signal after switching from outside air to filtered NO_2 free air.

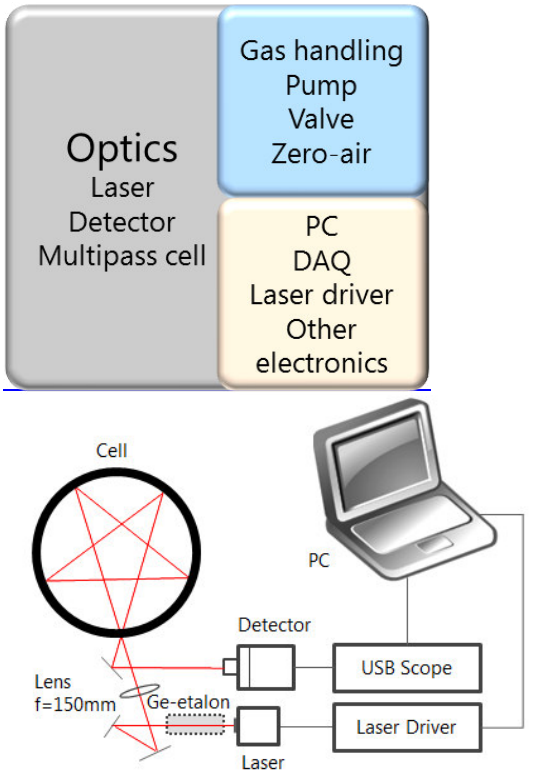
The detector signal is digitized by a digital oscilloscope (Picoscope 4000 Series, Pico Technology) at 12-bit resolution and with 20 MS/s sampling rate. The data acquisition is triggered by a TTL trigger signal generated by the laser driver. After acquisition of 1000 spectra at a rate of 2 kHz the data is transferred via a USB 2.0 port to a nano PC (Zero pro, Xi3 corporation) where the spectra are averaged and analyzed by a custom-written LabVIEW program. The NO_2 absorption spectrum is fitted with pressure, temperature, path length and line strength (HITRAN database (Rothman et al., 2013)) as input parameters to determine the NO_2 concentration according to Beer-Lambert's law. Pressure and path length are considered constant, while the temperature of the cell is measured with a thermistor and fed back to the fitting program. The chosen NO_2 absorption feature at 1599.9 cm^{-1} offers a large dynamic range from 0 to several ppm which is largely sufficient for ambient concentration measurements in cities.

Since the instrument is exposed to the temperature variations of the outside air without having active temperature control, we have to regularly compensate for instrument drifts. This is done on two levels: First, the set-point of the laser temperature is adjusted via a feedback loop to lock the laser frequency, and thus the NO_2 absorption feature is kept at the same position in the measured spectrum. This measure counteracts the instrument drift by regulating the laser temperature. Second, the instrument repeatedly determines the zero~~-~~ point offset, i.e. the instrument signal at zero NO_2, by measuring filtered NO_2-free air for 2 minutes every 20 minutes. This zero air is obtained by pumping outside air through a filter of 20 ml volume filled with Chemisorbant media (Purafil, Inc. USA).

Via a serial interface, the NO_2 concentration data (3 s averages) are transferred to a measurement unit of the "OpenSense" network. This unit is installed next to our instrument and described in detail by Hasenfratz et al. (Hasenfratz et al., 2015). The position of the tram is determined by GPS. NO_2 concentrations and GPS positions are transmitted via GSM (Global System for Mobile communications) and stored in a database.

(a)

(b)


Optics
Laser
Detector
Multipass cell
Gas handling
Pump
Valve
Zero-air
PC
DAQ
Laser driver
Other
electronics
Cell
PC
Detector
USB Scope
Lens
f=150mm
Ge-etalon
Laser Driver
Laser

(c)

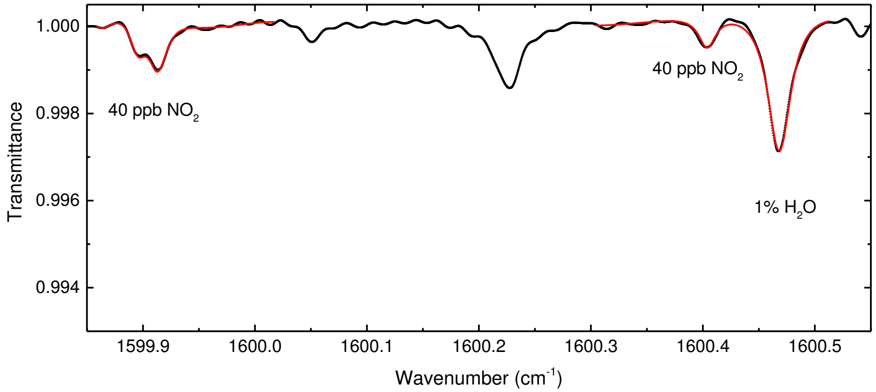

1.000
0.998
0.996
0.994
Transmittance
40 ppb NO$_2$
40 ppb NO$_2$
1% H$_2$O
1599.9
1600.0
1600.1
1600.2
1600.3
1600.4
1600.5
Wavenumber (cm^{-1})

Figure 1: (a) Photograph of the QCLAS instrument ~~with an optional display.~~without the cover. (b) Schematic of the ~~main components of the~~ instrument and its optical layout. (c) Absorption spectrum recorded by the instrument. The red lines show the areas that were fitted to determine concentrations.

2.2 Mobile NO_2 concentration measurements

The QCLAS instrument was placed on the roof of a tram of the public transport company of the city of Zurich (VBZ) from 21 Sept 2015 to 09 Mar 2016. This tram ran most of the time on the tram services no. 11 and 14 according to the operation schedule defined by VBZ (see supplementary materials for a detailed map). Tram services in Zurich operate from 5 a.m. to 1 a.m. at ~~altering~~varying intervals. Consequently, the set of daily measurements (time period, number and location) varies from day to day. The cross-city tram services no. 11 and 14 link areas between 400 and 520 m above sea level while elevated residential areas in Zurich are up to 640 m a.s.l. The routes of these services include short passages that are free of motorized traffic but mostly are on regular roads with little to heavy traffic.

We analyzed the NO_2 data from the period 09 Oct 2015 to 15 Feb 2016. Time periods for which high quality measurements are available are depicted in Figure 2. The gap between 27 Dec 2015 and 05 Feb 2016 is due to discarded data of reduced quality related to misalignment in the optical system of the instrument. Further, we excluded data when the tram was located within the depot area or when the GPS position was not clearly attributable to the correct tram track. Moreover, data ~~was~~were omitted when the NO_2 spectrum was not clearly identified by the processing algorithm. The water concentration determined from the measured spectrum was used as an indicator to flag such data. In total 37 % of the measurement days yielded data that could be used for statistical modeling.

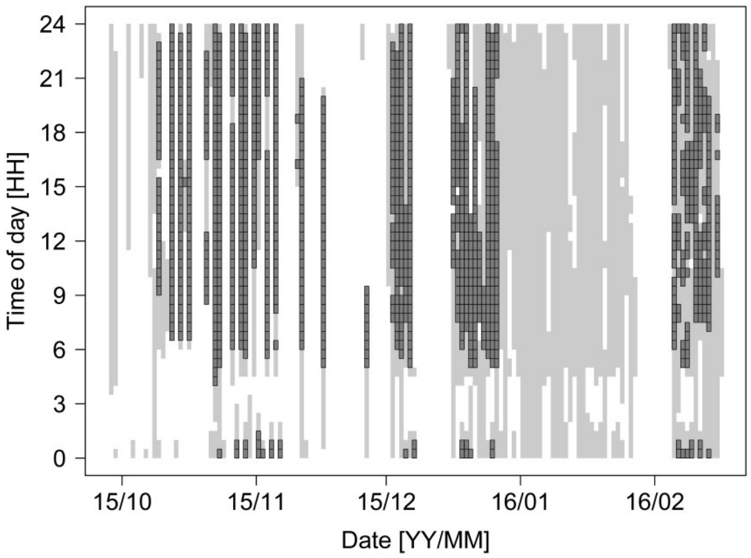


Figure ~~2:~~2: Time periods when the QCLAS instrument was in operation and NO_2 data were transmitted to the database are depicted in light gray. Thirty minutes periods when the QCLAS instrument obtained high quality NO_2 data and the tram was in regular operation are depicted in dark gray (total: 1183). Statistical models were developed for these periods.

2.3 NO_2 concentration measurements at fixed sites

The municipal (Department for Environment and Health Protection (UGZ), City of Zurich) and federal (Federal Office for the Environment, FOEN) authorities operate seven air quality monitoring (AQM) stations with CLDs for regulatory purposes in the city of Zurich. The set of stations includes background as well as highly polluted locations (see supplementary materials). It provides a comprehensive overview of the pollution situation encountered in Zurich with respect to the range of concentration levels and thus to the intra-urban variability. However, the spatial representation is coarse with respect to the generation of spatially highly resolved pollution maps. We had access to the one minute measurements from all sites for the time period 1 Jan 2015 to 1 June 2016. Two of the stations, STA and SCH, are located next to the tram tracks at a distance of 8 and 20 m, respectively. Five stations were operational during the complete campaign period. Operation of station BLU started on January 1, 2016. Station SWD was closed on February 1, 2016. The reported annual mean NO_2 concentrations in 2015 for the AQM sites were: HEU: 18 $\mu g/m^3$, ZUE: 31 $\mu g/m^3$, STA: 33 $\mu g/m^3$, SCH: 45 $\mu g/m^3$, SWD: 47 $\mu g/m^3$, RGS: 50 $\mu g/m^3$(OSTLUFT, 2016).

2.4 Spatial data

We employ spatial data for the classification of locations in the city of Zurich with respect to the instantaneous NO_2 concentration. NO_2 concentrations encountered in Zurich are moderate compared to other cities. Largest emission sources contributing to ambient NO_2 concentrations in Zurich are motorized traffic (47% of NO_x) and heating systems (28% of NO_x) (followed by industry (15%), constructions (9%) and others (1%)) (Brunner and Scheller, 2014). There is no heavy industry. Traffic emissions are by far dominant with respect to the spatial variability measured by our mobile instrument as the release points of the emissions from heating systems are usually well above street level.

We used average daily traffic volumes on particular roads, the digital elevation model, and building footprints and heights as spatial input data. The data was provided by the City and the Canton of Zurich and origins from the year 2013 which is well representative for the period observed in this study. From this data, we computed spatial variables representing the traffic intensity (several types), the "sky view factors" and the elevation above sea level following the procedures outlined in detail in Mueller et al., 2015 (Mueller et al., 2015) and Mueller et al., 2016 (Mueller et al., 2016). These variables cover the municipal area of Zurich in a 10 m grid.

Traffic intensity at a specific location was computed by summing up the distance travelled by vehicles in the ~~environment~~vicinity of this location. The summands were weighted based on an exponential decay function which depends on the distance from the location. Moreover, we multiplied heavy vehicles by a factor of 10 in order to account for the higher emissions compared to light vehicles. We used four types of traffic intensities: the first attributes a weight of 1/e to summands at a distance of 30 meters (DTV030), the second at a distance of 50 meters (DTV050), the third at a distance of 100 meters (DTV100) and the fourth at a distance of 400 meters (DTV400). While the versions DTV030 and DTV050 depict the near field traffic intensity, versions DTV100 and DTV400 account for the far field traffic intensity.

The variable "sky view factor" (SVF) indicates the fraction of the sky that is visible at a particular location by means of the surface area of half of a unit sphere being related to the built environment.

The variable type elevation (DTM) indicates the altitude above sea level of the measurement locations.

2.5 Statistical modelling

Statistical models directly rely on measurements. Accordingly, the distribution and accuracy of the measurements strongly impact the model results. The QCLAS NO_2 instrument installed on top of a tram provides an average NO_2 concentration for every three seconds. While the tram is moving, numerous different locations in the city are probed. These measurements are complemented by those from the fixed AQM sites. The combined data set allows the investigation of two aspects that are important for statistical modeling:

First, we focused on small-scale features of the pollutant field. This subject was addressed by computing models for the prediction of the NO_2 concentration at the locations of the moving tram. Models relying on measurements from AQM sites allow quantifying how well pollutant concentrations at different places in the city can be predicted based on the existing infrastructure consisting of seven fixed locations. Models based on mobile measurements shed light on the potential of denser measurement networks such as mobile ones in cities like Zurich. Second, we analyzed for which parts of Zurich the tram measurements provide information about the instantaneous pollutant field. Reference measurements for model validation are obtained from the spatially well-distributed AQM sites. Moreover, the application of the same modeling approach as used for the tram measurements to data from the AQM sites yields a benchmark for the performance of the mobile sensor network.

As outlined above, the modelling efforts are motivated by the request for accurate pollutant concentration maps. This study does not allow the generation of a long-term series of maps but adds knowledge of their generation.

An approach based on regression trees was developed~~, an approach that~~ (Breiman et al., 1984) and applied. Our approach relies on spatial information and NO_2 measurements and the basic assumption that NO_2 observations in the urban environment are similar at locations with similar spatial features. The main strengths of such a statistical modelling approach are the moderate computational time and the fact that information about the impact of instantaneous activities of emission sources and of meteorology is implicitly covered by the measurements.

The use of regression trees for the spatio-temporal analysis of the mobile and fixed NO_2 observations requires that any location in the city is described in terms of multiple spatial features. The chosen spatial features are DTV030, DTV050, DTV100, DTV400, DTM and SVF forming the location dependent input variables. The NO_2 observations and the input variables are linked by the GPS positions. Regression trees are described in detail in (Breiman et al., 1984). The basis of the approach is a set of data units referring to an observation y (mobile or fixed NO_2 measurement at location L_i on time t_i) and related values of the input variables j (i.e. the explanatory variables). The algorithm performs repetitive partitioning of the data set by means of the spatial features and the NO_2 measurements resulting in M location types, R_1, R_2, …, R_M. Figuratively, starting from the undivided data set a tree is grown that consists of splits and leaves (in our context: location types). Splitting variables j and split points s are found by aiming at reducing the variance of the NO_2 measurements referring to a location type by successively solving

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-c_2)^2\right]$$

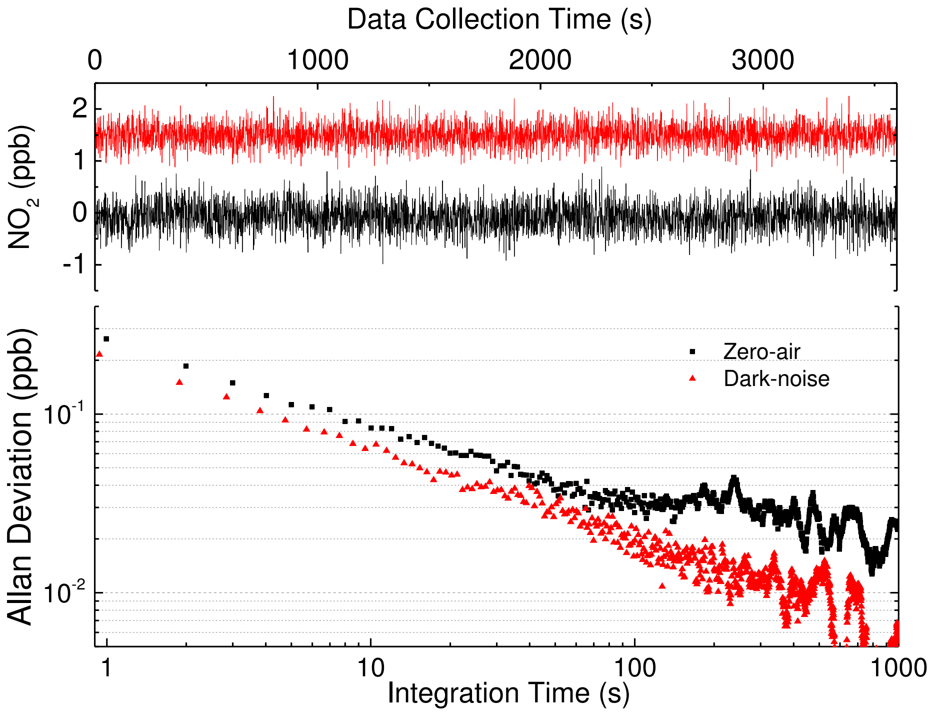
The modeled NO_2 concentration c_i of location type R_i equals the average of the NO_2 concentration measurements being part of R_i. The final number of location types or leaves depends on the number of successful splits. Splits are only attempted if at least 75 observations exist in a location type and the split increases the explained variance by 0.01. A location type R_i must contain at least 25 observations.

A location type R_i is the result of the splits in the domain of the spatial features. Accordingly, it does not represent a contiguous spatial area in reality but patches distributed in the city. NO_2 concentrations can be predicted for any location in the city based on the spatial features of these locations and the regression tree. We used the R package "rpart" (Therneau et al., 2015) for the computations.

3 Results ~~und~~and discussion

3.1 Instrument performance and stability

The precision of the instrument was determined using the Allan-variance technique (Werle et al., 1993). Figure 3 shows an Allan-deviation plot for measurements of filtered (NO_2-free) air in the laboratory. The best precision of 30 ppt is reached after ~~100~~200 s averaging while the 1 s precision is about 300 ppt. To get some insight into the source of the instrument noise, we performed a dark-noise measurement. In such an experiment the spectral fitting algorithm is applied to the detector output while the laser light is blocked. The resulting Allan-deviation for dark-noise is nearly equivalent to the Allan-deviation of the zero-air measurement, indicating that the NO_2 precision is basically limited by the dark-noise of the detector until the minimum is reached. While the Allan-plot for zero-air levels off at about ~~100~~200 s, due to instrument drifts, the noise of concentrations retrieved from dark~~-~~-noise can be further reduced by longer integration. Such behavior is expected ~~for~~in case of the white noise limited regime. The Allan-deviation of zero-air measurements shows the excellent stability of the instrument as it remains at a low level even after 1000 seconds. Based on these results, the measurement strategy was designed such that it applies a drift correction to the concentration measurements by measuring NO_2-free air for 2 min after 18 min of measuring. This offset-correction is achieved by subtracting the linear interpolation of the mean of the zero-air measurements before and after each 18 min measurement period.


Data Collection Time (s)
0
1000
2000
3000
NO_2 (ppb)
2
1
0
-1
Allan Deviation (ppb)
10^{-1}
10^{-2}
Zero-air
Dark-noise
1
10
100
1000
Integration Time (s)

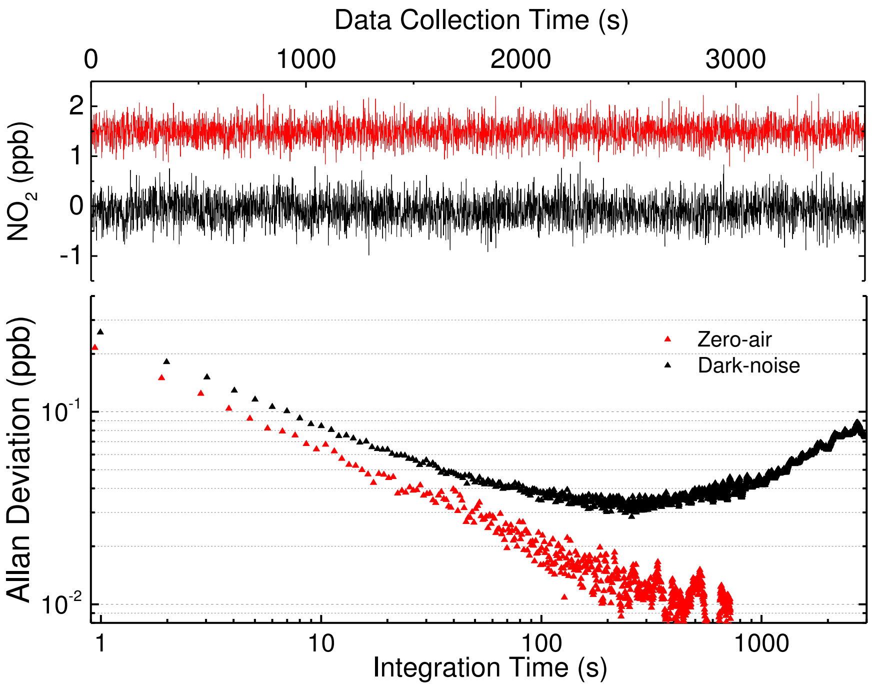


Figure ~~3:~~3: Determination of the instrument precision and stability: Time series of zero-air (black) and detector dark-noise (red) and the associated Allan-deviation plots. The dark-noise trace is offset by 1.5 ppb from zero for clarity.

The instrument accuracy was determined by comparing its retrieved values with data from the chemiluminescence detector (CLD~~) of a fixed air quality monitoring site (NABEL, located in Dübendorf, Switzerland).~~, APNA 370, Horiba) of a fixed air quality monitoring site (NABEL, located in Dübendorf, Switzerland). The CLD has a limit of detection of 0.1 ppb and at 30 ppb a total uncertainty of 1.25 ppb (NABEL, 2016). For this comparison, the instrument was installed on the roof of the air quality monitoring site next to the gas inlet of the CLD. A correction factor of 1.2 to the QCLAS data was found~~.~~ by plotting the QCLAS data over the corresponding CLD data. Such a correction to the normally calibration free QCLAS method was necessary because of loss of NO_2 during the sampling and in the long path cell and therefore was applied to all concentrations measured with the QCLAS. The contact of NO_2 with metal or plastic surfaces leads to dissociation. Since the pressure and flow rate in the cell are constant, the fraction of dissociated NO_2 can also be assumed as constant. Note that formation of NO_2 in the sampling line of instruments due to reaction of NO with O_3 can lead to an overestimation of NO_2. The influence of this artefact is small, especially because the measurements have been done during the cold season at low O_3 concentrations. It is, however, for the QCLAS a bit higher than for the reference CLD instrument, because the residence time of the air sample is for the CLD somewhat shorter (about 6s).

The concentration measurements of both instruments after scaling of the QCLAS data are depicted in Figure 4. They yield excellent agreement over the full range of concentrations that was encountered during the 48 h period that is displayed

(17 June 2015 to 19 June 2015). Discrepancies between the two instruments are mostly due to the higher temporal resolution of the QCLAS~~ and imperfect time synchronization.~~. The standard deviation of the QCLAS measurements from the CLD was found to be 0.96 ppb therefore we conclude that the instrument accuracy is about 1 ppb with respect to the CLD and that the assumption of a constant relative loss of NO_2 was correct.

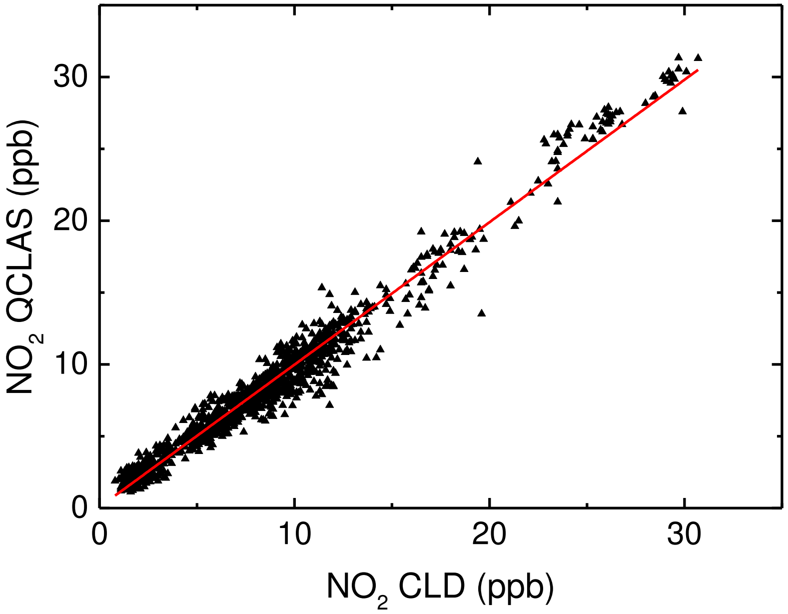


Figure ~~4:~~4: Scatter plot of QCLAS concentration measurements (1 minute means) versus CLD measurements (1 minute means). Both instruments ran in parallel at the AQM site in Duebendorf, Switzerland, in the period 17 Jun 2015 to 19 Jun 2015. The red line is a linear fit to the data with slope 0.99±0.005, intersect 0.08±0.05 and an R^2 of 0.97.

3.1.1 GPS positioning and map matching

The measuring unit on the tram was equipped with a GPS receiver providing a position every ~~third second.~~three seconds. GPS positions were orthogonally projected on the nearest tram track after the route of the tram had been determined. We discarded all the positions with corrections exceeding 10 m.

NO_2 measurements and GPS positions were not synchronous. Positions of the NO_2 measurements were linearly interpolated from the matched GPS positions. Hereby, we added ~~11.~~1.5 seconds to the timestamps of the matched positions~~ in order to account for~~. Such a delay is necessary because at a given time-stamp the ~~flushing time of ~~GPS provided the ~~measuring cell and for half of~~present position while the ~~length of~~NO_2 concentration is the ~~measurement interval~~average of 3 seconds before a time-stamp.

The NO_2 measurements ~~refer to~~represent 3~~ seconds averaging time. Therefore, they are strictly speaking not point measurements but refer to~~ s averaged data, which, considering the traveling speed of the tram, cover a given route segment instead of representing a fixed point measurement. The length of this segment is the product of the integration time of 3 seconds and the speed of the tram. Tram speed associated with the mobile NO_2 measurements was smaller than 5 m/s for 64 % and smaller than 10 m/s for 92 % of the time.

3.1.2 Characteristics of tram based NO_2 measurements

Operation conditions during the mobile application are significantly more demanding than in the laboratory due to the harsh environmental conditions involving large temperature variations, vibrations, changing humidity and ~~rain.~~precipitation. We analyzed the quality of the data obtained in the mobile application ~~twofold~~in two ways: First, we quantified the measurement uncertainty related to noise and variations of the zero point offset. Second, we compared the tram measurements to measurements from a fixed air quality monitoring site (see next section).

Figure 5 (a) depicts the NO_2 zero point offsets determined during the 2 minutes periods when NO_2-free air is pumped into the measuring cell (the cell is flushed during the first 45 s of the 2 minutes period; 25 measurements (75 s) remain to determine the zero point offset). NO_2 offsets outside the range [-10..10] ppb and NO_2 offsets derived in zeroing periods with less than 5 measurements or with a standard deviation exceeding 5 ppb were discarded. ~~87~~88 % of the remaining zero point offsets are within ± 5 ppb. The corresponding standard deviation of a single measurement during the zeroing period is in the order of 0.5 ppb (Figure 5 (b)). Measurements that are not enveloped by two zeroing periods ~~in~~within a time interval of less than 20 minutes cannot accurately be adjusted for the zero point offset and were ~~omitted~~rejected.

We corrected the NO_2 measurements by linearly interpolating the zero point offsets derived in two consecutive zeroing periods. An upper limit for the interpolation error was derived by predicting the zero point offset of a 2 minute zeroing period by linear interpolation of the two neighboring zero point offsets (40 minutes time span). The resulting standard deviation of the difference between predicted and measured zero point offset amounts to 1.5 ppb (Figure 5 (c)). Zero point offsets are temporally correlated as the standard deviation of the differences of two consecutive zero point offsets amounts to 2.1 ppb and thus is smaller than 4.3 ppb expected for white noise ($\sqrt{2} \cdot \sigma_{NO2\ ZPO}$).

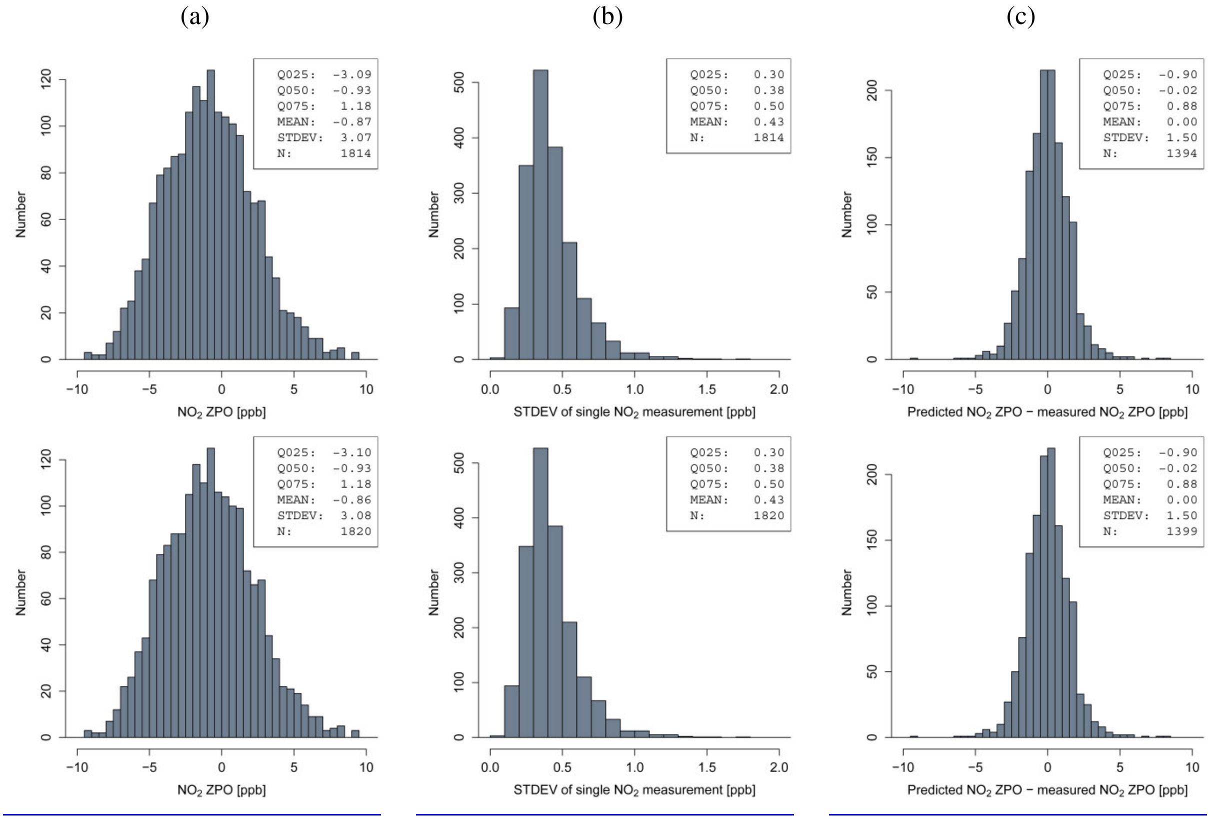


Figure ~~5:~~5: (a) Histogram of NO_2 zero point offsets (ZPO) computed during the calibration periods. Q025, Q050 and Q075 denote the 25%, 50% and 75% quantiles, respectively. (b) Histogram of the standard deviations of a single measurement during the zero point measurements. (c) Differences between predicted and measured zero point offsets.

3.2 Intra-urban and temporal variation in NO_2 concentration

The measurements of the mobile QCLAS instrument reveal the spatio-temporal variation of the urban pollutant concentration field. As an example, Figure 6 shows one day time series of tram measurements (3 seconds and 5 minutes averages) and of measurements (5 minute averages) from the fixed monitoring stations SCH and STA. The fixed air quality monitoring stations STA and SCH are located at 8 and 20 m distance from the tram tracks, respectively, and are equipped with CLDs measuring NO_2. The tram passed the fixed sites about every 30 minutes on the selected day. Site SCH is impacted by heavier traffic than site STA resulting in higher NO_2 concentrations on average. Annual mean concentrations at SCH and STA in 2015 were 23.6 ppb and 17.3 ppb. Traffic in the morning and evening rush hours in Zurich ~~are~~is comparable. Hence, a large part of the diurnal variation in NO_2 concentration is due to meteorology, e.g. the development of the boundary layer height during the course of the day.

The tram carrying the QCLAS instrument regularly passed AQM sites STA and SCH during the measurement campaign. This provides the opportunity to directly compare the measurements from the tram (3 seconds averages) and the measurements from the fixed site (1 minute averages). We plotted single measurements from the mobile instrument that were obtained in a 30 m radius around the fixed station STA (a) and SCH (b) over the corresponding 1 min average values from these AQM stations in Figure 7. The comparison reveals excellent agreement of measurements from the QCLAS instrument to measurements from site STA, whereas the agreement between the tram measurements and the measurements from station SCH is ~~not as good.~~less pronounced. This can be expected for two reasons: (i) the larger distance, (ii) the location of site SCH at a crossroad where traffic flow is controlled by traffic lights~~. The calibration~~ and where the short term variability of ~~sensors using precise measurements~~NO_2 is larger. From a more general perspective, this is of importance when mobile sensor data is corrected based on data from ~~AQM~~nearby fixed reference sites~~ is an option for mobile operated sensors~~, as suggested by (Arfire et al., 2015; Saukh et al., 2015). ~~However, this analysis clearly shows the advantage of a mobile measurement device that does not rely on fixed sites for~~The parameters for data correction as determined from such an approach depend on the ~~determination~~local concentration variability and the time response of ~~its calibration parameters.~~the sensor and the reference instrument. Probing different air parcels may concur with concentration differences as it is observed for the data set obtained for AQM site SCH. Parameters determined from such a data set could be associated with significant errors depending on the temporal stability of the sensor, the applied mathematical sensor model and the term used in the mathematical model for the description of the concentration differences.

Figure 8 depicts the NO_2 measurements on the track section from station “Bellevue” to “Bahnhofquai” ~~aggregated by~~for the full measurement campaign (148 runs) with 50 ~~meters of track length. It~~meter resolution. Figure 8 shows the small-scale NO_2 variations that are measured by the mobile instrument. At sections where the tram is operating on congested roads (between the first and the third stop as well as between the fifth and the seventh stop), the average NO2 concentration is higher than at sections with very little traffic.Moreover, ~~it~~Figure 8 shows the spatial distribution of the number of measurements taken, which is highest close to tram stations where the velocity of the tram is ~~reduced~~low or zero.

Our statistical models are driven by measurements and are based on the assumption of comparable NO_2 concentrations at ~~similar locations.~~locations that are similar in terms of vicinity to sources (in particular traffic), the built environment and elevation. They cannot provide accurate predictions for locations with NO_2 concentrations significantly higher or lower than observed. We analyzed this subject by a comparison of all the compiled tram measurements and the measurements from each AQM site (5 minutes averages, respectively). Table 1 shows that NO_2 concentrations measured on the tram are in line with concentrations measured at AQM sites (roadside, urban background) in the city center. They are most similar to the measurements from the AQM site STA (RMSE: 7.8 ppb, slope: 1.10, Δ mean=3.2 ppb). Site STA is most representative for the set of location types passed by the tram. In contrast, NO_2 concentrations at the elevated (200 m above the city center) background site HEU are rarely in the range of the tram measurements, demonstrating the limited representativeness of the tram-based data set for the entire city.

Table 1: Comparison of NO_2 tram measurements and measurements from the AQM sites, aggregated to 5 minutes means. N denotes the number of ~~values~~measurements, r is the Pearson correlation coefficient, and slope refers to a regression line through the origin (tram against AQM), Δ mean is the difference between the average concentrations (tram minus AQM) and Q005, Q050 and Q095 denote the 5%, 50% and 95% quantiles of the differences tram minus AQM.

AQM site	N	RMSE [ppb]	r	slope	Δ mean [ppb]	Q005 [ppb]	Q050 [ppb]	Q095 [ppb]
BLU	1148	11.6	0.72	1.33	8.0	-4.3	7.1	22.~~4~~3
HEU	~~6606~~6608	15.4	0.58	1.65	12.4	0.4	10.9	29.3
RGS	~~6609~~6611	10.4	0.~~58~~57	0.86	-3.4	-20.9	-2.4	10.3
SCH	6220	9.5	0.67	0.88	-2.4	-17.~~7~~6	-2.0	11.~~6~~5
STA	~~6647~~6649	7.8	0.76	1.10	3.2	-7.3	2.7	14.8
SWD	~~5468~~5470	11.3	0.46	0.85	-3.1	-20.7	-2.8	13.9
ZUE	~~6556~~6558	9.5	0.72	1.22	5.6	-5.2	4.~~8~~9	18.~~6~~5

The temporally highly resolved tram-based measurements provide information about the NO_2 concentration distribution along the track. We analyzed the distribution of the NO_2 tram measurements in 30 minutes modelling periods (Ti) by computing the 5% and the 95% quantile of the measurements as well as the difference between the 95% and the 5% quantiles (Figure 9 (a) and (b)). The difference between the 5% and the 95% quantiles, respectively, and the mean value remains rather constant for increasing mean values. This points out that temporal variation (i.e. the mean values of different modeling periods) in the tram based NO_2 measurements is of comparable magnitude to intra-urban variation (difference between the 95% and the 5% quantiles within a modeling period). The difference between the 95% and the 5% quantiles is below 35 ppb in about 90% of the modelling periods. Short-term variation in NO_2 concentration at a specific location is mainly caused by changing activities of emission sources in the close vicinity. This is observed at all AQM sites in Zurich that are impacted by traffic (Figure 9 (c) and (d)) and can be of similar magnitude than the NO_2 variations measured by the tram in motion.

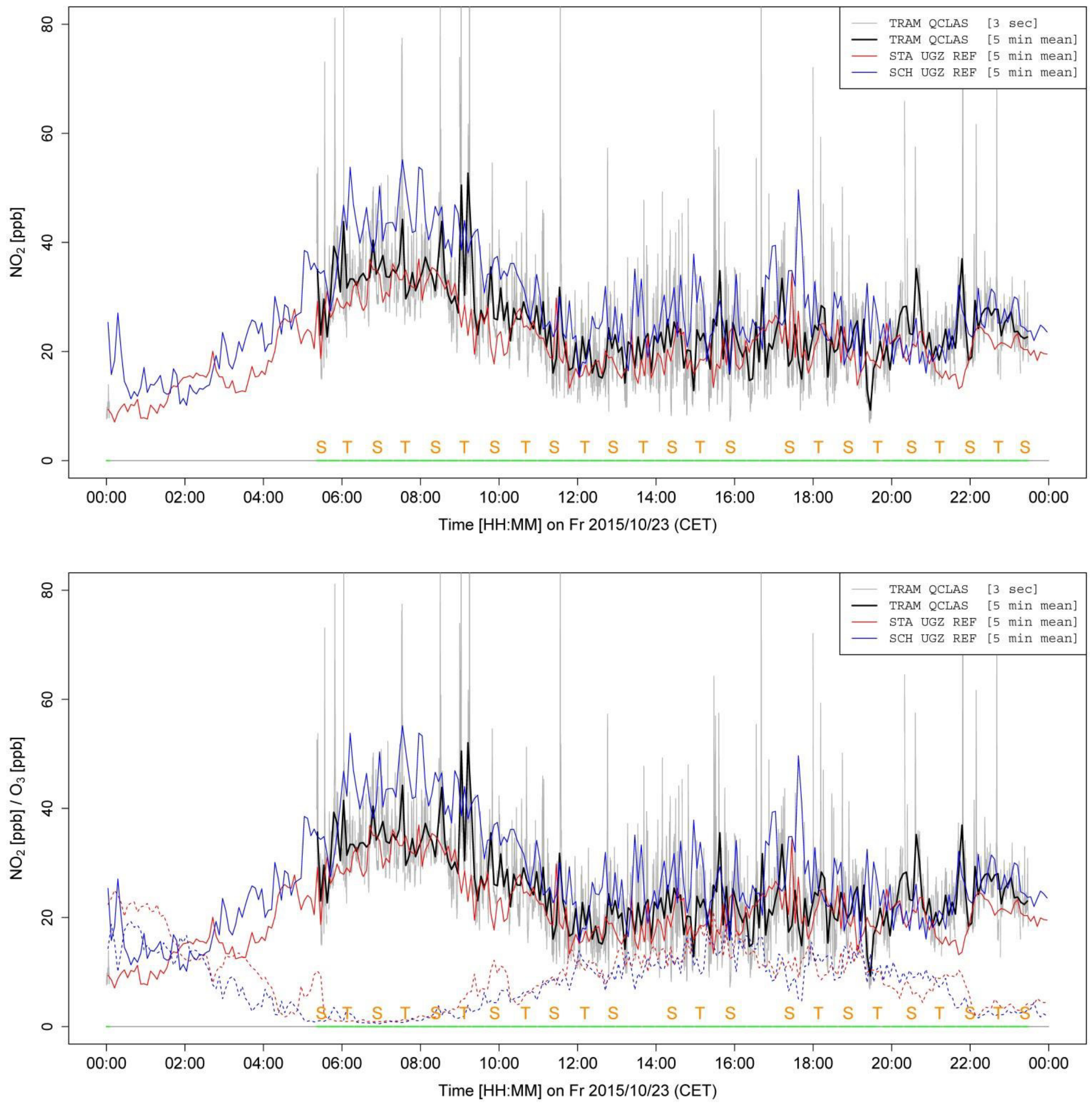


Figure ~~6:~~6: Example time series of measurements of the tram based QCLAS instrument (single measurements and 5 minutes means) as well as of NO$_2$ and O$_3$ measurements from the AQM stations STA and SCH (5 minutes means). The letters "S" and "T" depict when the tram passes the terminal stations Seebach and Triemli.

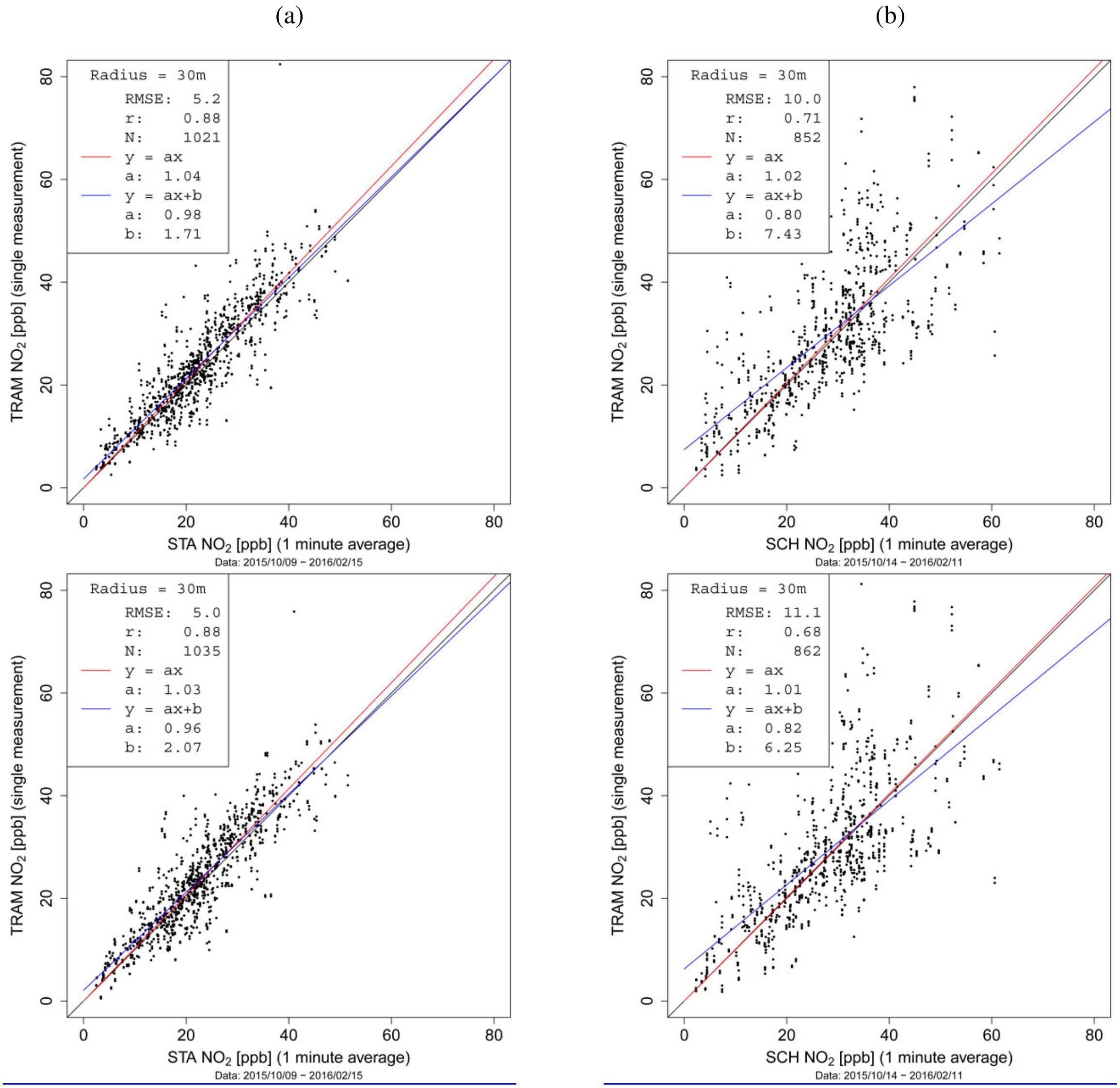


Figure ~~7:~~7: Comparison of measurements from the QCLAS instrument (single measurements) and measurements (1 minutes mean) from the AQM sites STA (a) and SCH (b) when the tram is within a 30 meter radius from the respective AQM site. RMSE is the root mean square, r denotes the Pearson correlation coefficient and N is the number of 3 seconds measurements from the tram.

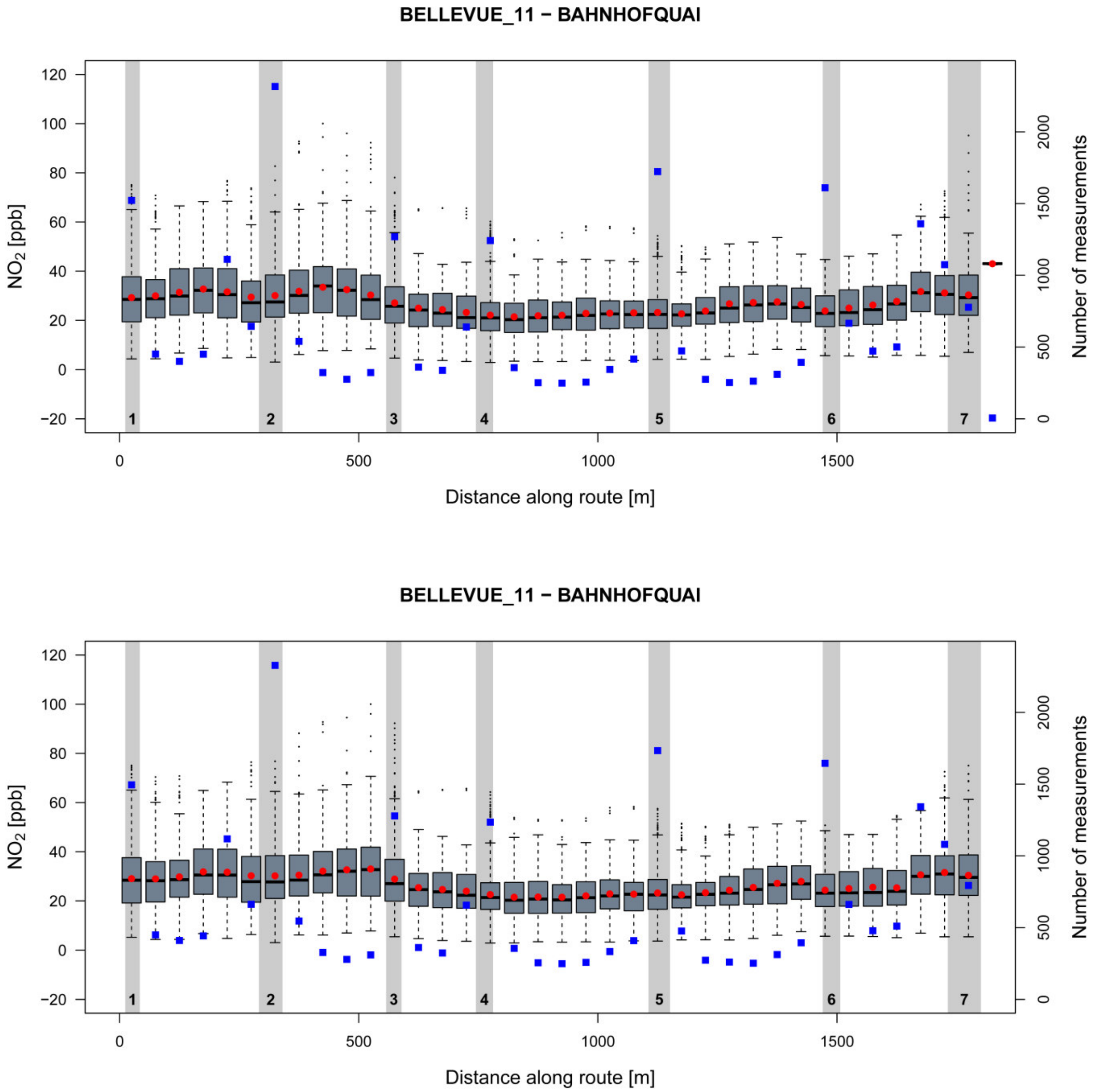
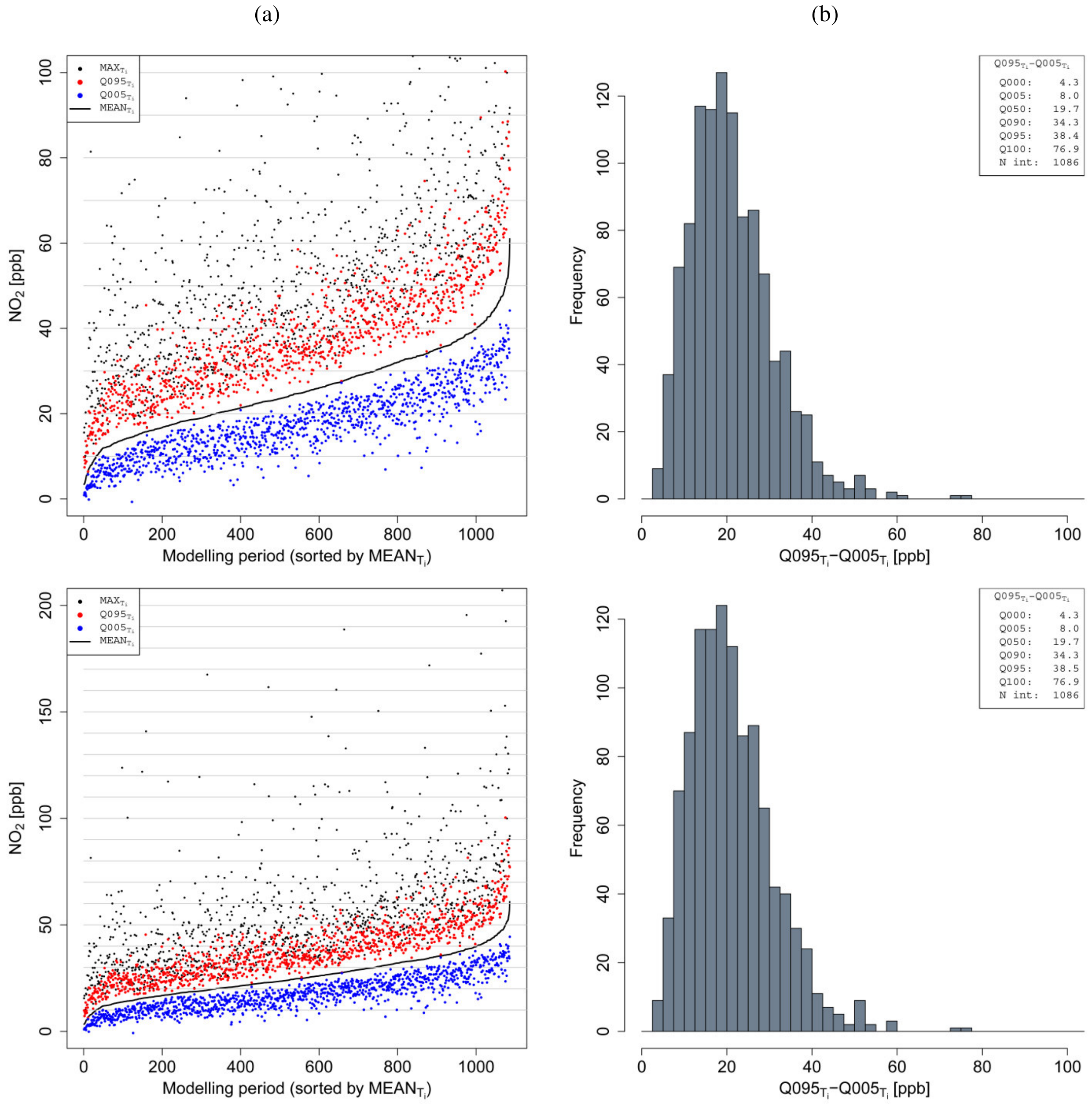


Figure ~~8:~~ 8: Boxplot of QCLAS measurements taken on all runs (n=148) between station “Bellevue” and station “Bahnhofquai” during the entire campaign~~. They are depicted in boxplot style after applying an aggregation of 50 meters~~ and to 50m track length. The red dots ~~depict~~indicate the NO_2 concentration means of the boxes, the blue rectangles depict the number of measurements referring to particular boxes~~.~~, the black dots are measurements outside the whiskers. Most measurements were taken near tram stops (depicted in gray, black numbers). The tram is directly impacted by traffic between the first and the third stop as well as between the fifth and the seventh stop. There is only little traffic between the third and the fifth stop.


(a)
(b)
MAX_{T_i}
$Q095_{T_i}$
$Q005_{T_i}$
$MEAN_{T_i}$
NO_2 [ppb]
0
20
40
60
80
100
Modelling period (sorted by $MEAN_{T_i}$)
0
200
400
600
800
1000
$Q095_{T_i}-Q005_{T_i}$
Q000: 4.3
Q005: 8.0
Q050: 19.7
Q090: 34.3
Q095: 38.4
Q100: 76.9
N int: 1086
Frequency
0
20
40
60
80
100
120
$Q095_{T_i}-Q005_{T_i}$ [ppb]
0
20
40
60
80
100
MAX_{T_i}
$Q095_{T_i}$
$Q005_{T_i}$
$MEAN_{T_i}$
NO_2 [ppb]
0
50
100
150
200
Modelling period (sorted by $MEAN_{T_i}$)
0
200
400
600
800
1000
$Q095_{T_i}-Q005_{T_i}$
Q000: 4.3
Q005: 8.0
Q050: 19.7
Q090: 34.3
Q095: 38.5
Q100: 76.9
N int: 1086
Frequency
0
20
40
60
80
100
120
$Q095_{T_i}-Q005_{T_i}$ [ppb]
0
20
40
60
80
100

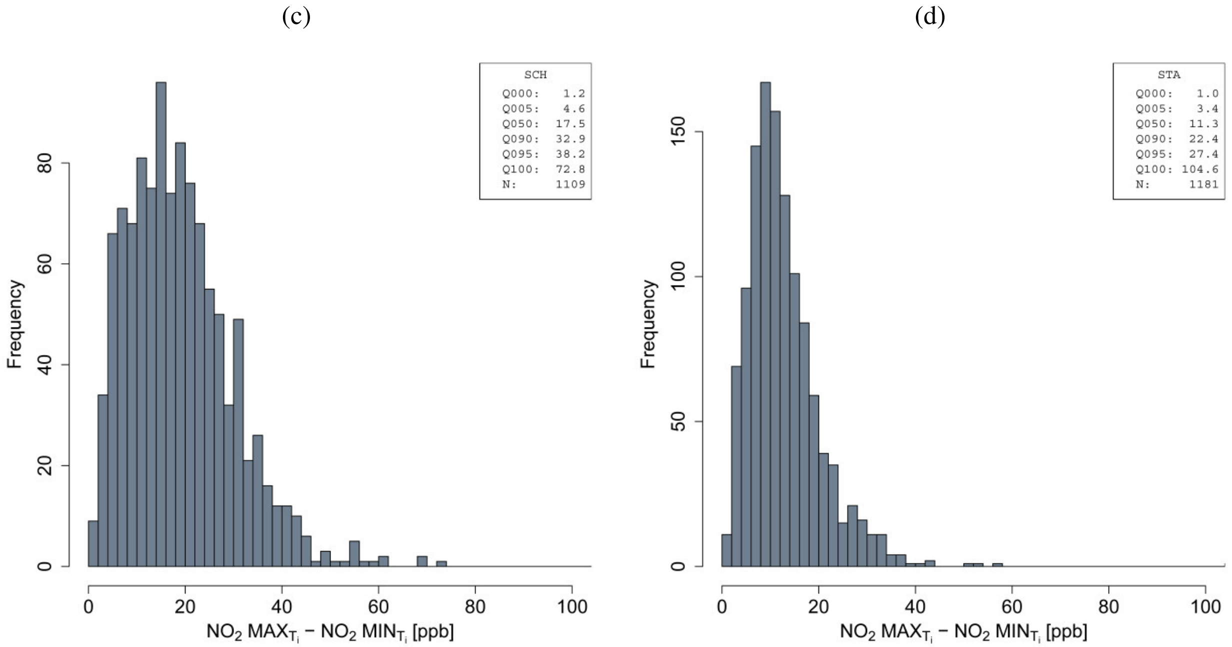


Figure ~~9:~~9: (a) 5% quantile (Q005$_{Ti}$; blue dots) and 95% quantile (Q095$_{Ti}$; red dots) of the NO$_2$ measurements (3 seconds) as well as the maximum (black dots) and the mean (black line) in each modelling period T_i (30 minutes duration, approximately 500 measurements; periods T_i with less than 300 measurements were not used for the plots in (a) and (b)). The intervals are ordered by the mean. (b) Distribution of the differences between the 5% and the 95% quantiles of the measurements in the modelling periods T_i. (c) Histogram of the differences between the maximum and the minimum 1 minute NO$_2$ measurements from AQM site SCH in the modelling periods T_i (30 minutes). (d) Same as in (c) for AQM site STA.

3.3 Results of statistical modelling

We applied the procedure described in section 2.5 to selected sets of NO$_2$ measurements obtained by the mobile QCLAS instrument and the fixed AQM stations. Independent regression trees are computed for each 30 minutes period. Setting the modelling period to 30 minutes is a compromise between the number and spatial distribution of mobile measurements forming the input for the regression tree computation and the temporal variation of the NO$_2$ concentration, i.e. changes in emission source activity and meteorology which may alter the relation between covariate values and NO$_2$ concentration.

The locations within the municipality of Zurich were described by means of six spatial variables (i.e. explanatory variables). Various combinations of these variables were defined, and regression trees relying on these variable selections were computed in order to find the best performing model for each particular setting during all modelling periods. ~~The frequency a particular variable is effectively used for splitting the 30 minutes data set depends only on the measurements.~~

The number of location types for which representative NO$_2$ concentrations can be determined based on our statistical models is limited. The number of measurements required for estimating an accurate mean concentration for a location type depends on the short-term temporal variability of the NO$_2$ concentration (see Figure 9 (c) and (d) and section 3.3.2). However, the number and spatial distribution of the measurements in a modelling period (approximately 500) ~~and the short-term variation of NO$_2$ concentration at a~~is given ~~location (see Figure 9 (c) and (d)) limit the possibilities for separation of different location~~

~~types with respect to the NO_2 concentration~~ by ~~a classification approach.~~the tram route and cannot be influenced. Accordingly, the use of too many explanatory variables in our statistical models may lead to an over-determined model. Regression trees based on tram measurements have about 6 location types R_i on average.

Models based on a single variable perform best when relying on a traffic related variable. Thus, elevation and sky view factors are of secondary importance. Predictions based on a basic model can easily be validated by cross-validation and by comparison ~~to~~with measurements from fixed AQM sites.

3.3.1 Model validation by means of three seconds NO_2 measurements from the tram

Three seconds mean NO_2 concentration referring to time and location of the tram were predicted in three versions. In version V_T, the data of the 30 minutes interval were partitioned into six five minutes intervals, denoted as A to F. Three independent regression trees were determined based on the tram measurements from BCEF, ACDF and ABDE and model predictions for AD, BE and CF were computed~~.~~, respectively. In version V_{AQ}, regression trees based on the measurements from the fixed air quality monitoring sites were computed. In version V_{STA}, 30 minutes means of the measurements from the fixed site STA were taken as predictions for the tram measurements.

The tram measurements as well as the computed predictions were aggregated to 5 minutes averages for being used in the statistical analysis. The predictions based on the tram measurements (version V_T) are best followed by those based on measurements from the fixed sites (version V_{AQ}) and those based on measurements of site STA (version V_{STA}). The difference in performance between V_{AQ} and V_{STA} is small due to the similarity of NO_2 concentration encountered at site STA and by the tram-based instrument. The results are summarized in Table 2.

Table 2: Summary of the comparison of model predictions and measurements from the tram (5 minutes means, respectively). r is the Pearson correlation coefficient, slope denotes a regression line through the origin (predictions against observations), N denotes the number of 5 minutes values. The variables used in the relevant model are listed in the last column.

Setting	RMSE	r	Slope	N	Used variables
V_T	5.~~9~~6	0.~~85~~86	0.97	~~6573~~6584	DTV050
V_{AQ}	7.3	0.~~77~~78	0.89	~~6604~~6599	DTV030/DTM
V_{STA}	7.6	0.79	0.84	~~6592~~6588	-

3.3.2 Model validation by means of 30 minutes mean NO_2 concentrations at fixed AQM stations

We computed a set of regression trees ~~utilizing~~using different sets of measurements for each 30 minutes modelling period in order to predict NO_2 concentrations at the AQM sites. Version W_T is based on tram measurements only, version W_{AQ} is based on the measurements from the fixed AQM sites only (excluding the measurements from the target AQM site, respectively), and version W_{T+A} is based on the tram measurements and the measurements from the fixed AQM sites (excluding the measurements from the target site). We attributed a weight of 20 to the AQM site measurements with respect to the QCLAS measurements in version W_{T+A}. Models of version W_T rely on the variable DTV100, models of versions W_{AQ} and W_{A+T} rely on the variable DTV030. Model versions ~~utilizing~~using these variables provide the best overall fit to the measurements of all AQM sites. In version W_B we simply use the 30 minutes mean NO_2 concentration measured by the tram as prediction for the concentration at respective AQM sites as benchmark for versions W_T, W_{AQ} and W_{T+A}.

A series of three maps illustrates differences between the versions V_{AQ}, V_T and V_{T+A} in terms of the measurement distribution and the resulting NO_2 concentration field (Figure). For optimal comparison, all these maps are based solely on the variable DTV050. Concentrations are shown only for locations with DTV050 values within the DTV050 range covered by the measurements as extrapolations based on statistical models are associated with larger uncertainties. Obviously, combining the measurements from AQM sites and from the tram yields the data set with the largest spatial coverage. The figure shows that the spatial coverage is not optimal, yet, in any version. Considerable potential remains in the optimization (i.e. extension) of static or mobile measurement networks.

Figure depicts a set of indicators for the agreement between model predictions (W_T, W_{AQ}, W_{T+A} and W_B) and 30 minutes mean concentrations measured at the AQM sites. The results show that the classification approach in version W_T reveals reasonable, bias free predictions for sites that are similar to locations passed by the tram (slope: 0.~~94~~93-1.~~17~~18, differences of average concentrations Δ_{mean}: -~~0.8~~1.1-4.~~7~~8 ppb). The scattering in version W_T is not smaller than in version W_B but this can partly be related to the partitioning of the tram measurement set into m location types R_i by the classification approach. ~~The~~

In our statistical models, the mean of all the measurements in R_i (covering approximately 1/m of 25 minutes with mostly m ~~=~~ = 6) is taken as the estimate value for the 30 ~~minutes~~minute mean concentration of location type R_i. We ~~assessed the scatter related to this kind of underlying data partitioning by a simple~~performed the following computation ~~based on the data of~~in order to analyze the agreement between the mean derived from a limited number of measurements and the true 30 minute mean concentration for several locations. For this analysis, measurements from single AQM sites from the year 2015~~.~~ ~~Thus, we~~ were used. We computed 5 ~~minutes~~minute means and complementary 25 ~~minutes~~minute means (6 pairs in 30 minutes, respectively). The obtained RMSE values range between 1.5 ppb at HEU (elevated background site) and 7.5 ppb at RGS (roadside). Therefore, differences between the mean of a subset of tram measurements referring to a particular location type and its 30 minute mean have to be expected if NO_2 concentration at this location is highly variable on the short-term.

The results of model version W_{AQ} are comparable to version W_T which points to a good spatial distribution of the monitoring stations. The combination of measurements from AQM sites and from the tram (Version W_{T+A}) performs best as this model version, in contrast to W_T, also includes information about the urban background.

The tram instrument is rarely situated in background locations. Hence, NO_2 concentrations are overestimated by W_T at AQM sites representing these types of locations (BLU, HEU and ZUE). To overcome this limitation, NO_2 concentration measured by the tram instrument might be considered as resulting from a background concentration plus an increment depending on the activity of nearby emission sources. Accordingly, urban background concentrations can also be estimated based on a baseline instead of the mean of a particular set of measurements. We defined the urban NO_2 concentration baseline as the 20% quantile of the NO_2 concentration measurements from the tram instrument in 30 minutes. Comparison of these quantile values with the means of the urban background site ZUE results in RMSE, Δmean and slope of 5.~~2~~1 ppb, 0.5 ppb and 1.00, respectively. The baseline approach therefore outperforms version W_T for site ZUE.

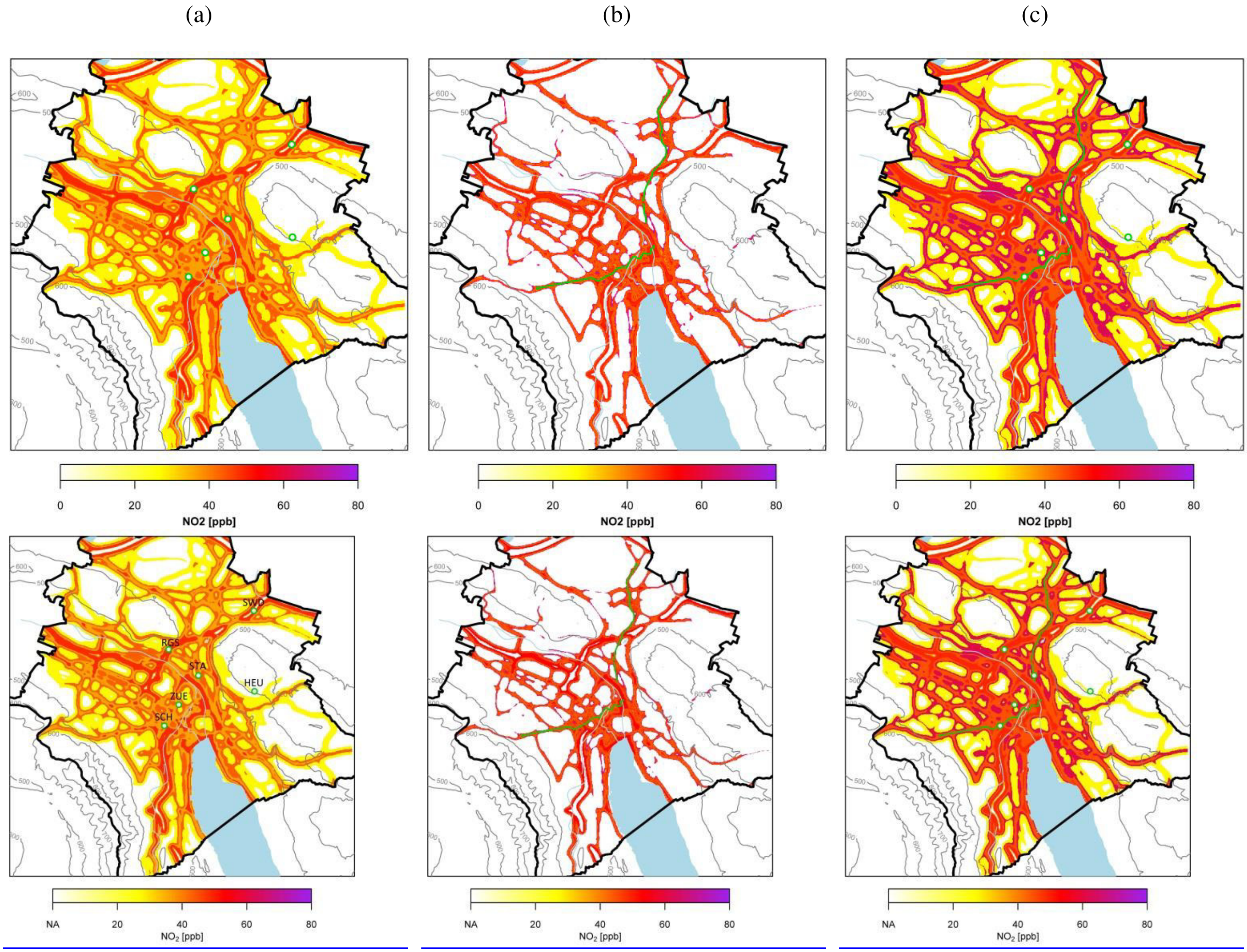


Figure 10: Series of NO_2 concentration maps for 17 Dec 2015 18:00-19:00. The models are based on the variable DTV050. Measurement locations are depicted by green dots, the tram track by a green line. (a) Map based on data from AQM sites (Version V_{AQ}). (b) Map based on measurements from the tram based QCLAS instrument (Version V_T). (c) Map based on measurements from the AQM sites and the QCLAS instrument (Version V_{T+A}). The maps cover a 12 × 12 km area. Light blue lines and areas depict rivers and lakes, the thick black line shows the border of the municipal area of Zurich and the gray lines indicate the contour lines of elevation (500 to 800 m a.s.l.).

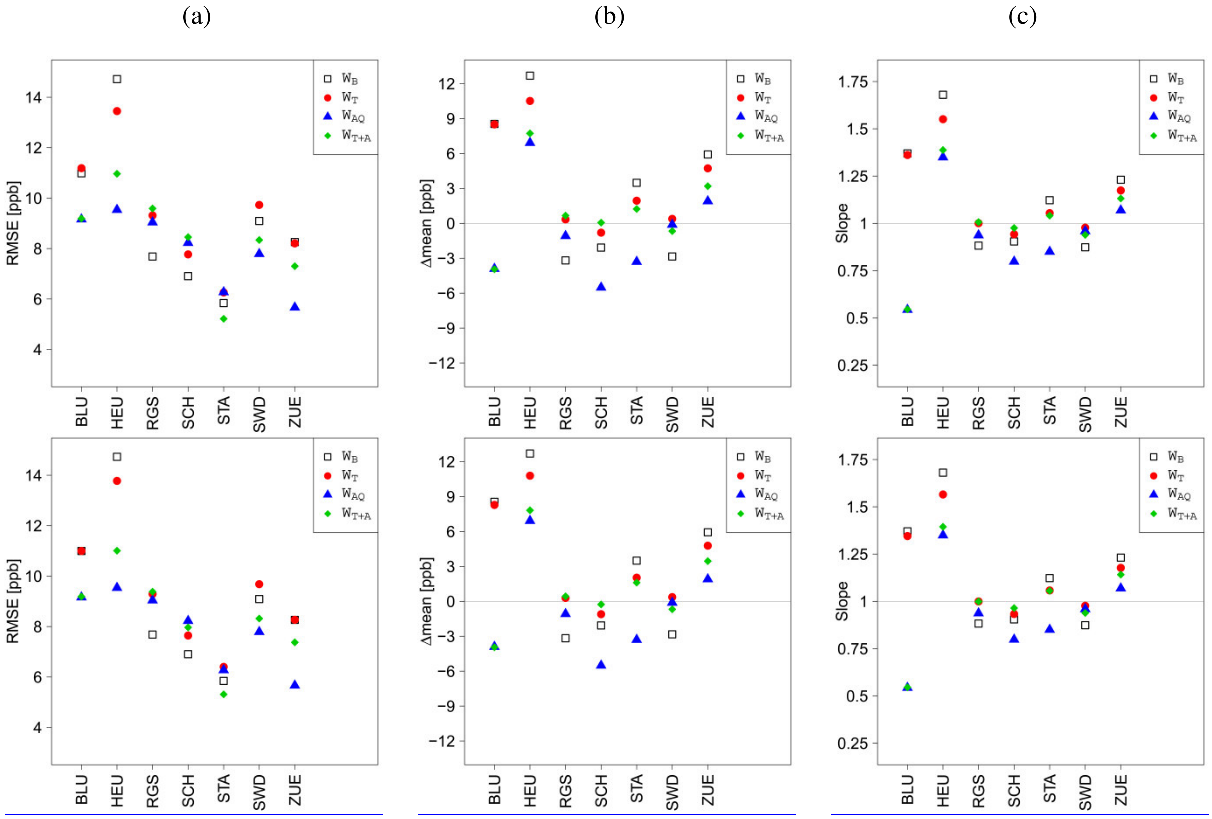


Figure 11: Comparison between observed 30 minutes means at AQM sites and model predictions based on the versions W_B, W_T, W_A and W_{T+A}. RMSE denotes the root mean square error, Δmean equals 1/N·(∑predictions - ∑observations) and slope refers to a best-fitting regression line through the origin (predictions over observations). N is 222 for site BLU, 955 for site SWD and between 1109 and 1183 for the other sites.

3.4 Discussion and outlook

The mobile QCLAS instrument provided accurate NO_2 measurements of high spatial resolution but with sparse sampling for particular location types related to the accessibility of the measurement locations that were limited to the tram tracks (few background locations, no elevated locations). In principle, this kind of concentration data set can be improved for statistical modelling by the operation of a larger number of instruments. Options are the installation of additional instruments on further trams and on other public transport vehicles (e.g. buses connecting residential areas running on roads with very low traffic) or the operation of instruments at fixed background sites or at highly polluted locations. ~~Our results~~Due to the number of required measurements for location types with highly variable NO_2 concentration, the necessity for measurements at locations with different pollutant situations and the travel time from one to another location, we suggest that the number of required mobile or static QCLAS instruments for comprehensive NO_2 measurement in Zurich is not smaller than the current number of fixed AQM sites.

The models derived from tram measurements yielded the most accurate predictions for tram measurements. This is related to the small spatial distance between measurements and predictions as well as to the dense spatial distribution of the tram measurements. The latter corroborates the potential of dense measurement networks, either static or mobile ones, for obtaining more detailed information on the instantaneous pollutant field.

Mobile measurement devices that provide accurate long-term measurements cannot only contribute to map generation but also to the validation of products derived from pollutant maps. ~~The model type the maps are based on is irrelevant in this context.~~ For example, the ~~accuracy of computed~~ exposure ~~values could~~estimations can be ~~quantified~~validated based on measurements of such instruments.

With the latest dual-wavelength QCL technology (Jagerska et al., 2014; Süess et al., 2016), the next generation of the QCLAS instrument could measure NO and NO$_2$, simultaneously~~ within~~, with the same ~~footprint. These~~compact geometry. Such measurements would give valuable insights into urban NO$_x$ chemistry. Furthermore, other target species, such as CO, CO$_2$, O$_3$ or CH$_4$, may in the near future be included.

4 Conclusions

We presented a compact and robust QCLAS spectrometer that operated autonomously on the roof of a tram to measure NO$_2$ concentrations in Zurich over a period of four months~~.~~ (with an interruption of five weeks). The analysis of the measurements from periods when the instrument was operated next to a CLD of a fixed AQM site and from mobile operation on a tram showed that the accuracy of 3 second values in the field is better than 2 ppb~~. Independency~~ with respect to the CLD that was used for initial calibration. Independence from roadside calibration infrastructure as well as from on-board calibration gases is a clear asset of the instrument for operation in urban environments with highly variable NO$_2$ concentration fields.

The QCLAS instrument is highly suited for mobile applications. The statistical modelling showed that the mobile measurements provided accurate information about the urban NO$_2$ concentration field. However, comprehensive analysis of the data from the tram and the AQM sites revealed large spatio-temporal variability in NO$_2$ concentration in the vicinity of emission sources. Accordingly, the main shortcomings of the obtained data set with respect to statistical modelling are the limited number of observations and the incomplete coverage of different types of locations. Improving the data set would require the operation of a larger number of instruments. This would allow the mapping of the entire NO$_2$ concentration field in a city with high spatio-temporal resolution and, concurrently, the refinement of statistical modelling techniques.

Acknowledgements

The support from the group of Computer Engineering & Networks Lab of ETH Zurich in the transfer and storage of the mobile NO$_2$ measurements is gratefully acknowledged. In this context, we thank Balz Maag for his help to install the instrument on the tram. The Department of Environment and Health Protection (UGZ) of the City of Zurich, the Department for Waste, Water, Energy and Air (AWEL) of the Canton of Zurich and the Swiss Federal Office for the Environment (FOEN) is acknowledged for providing NO$_2$ concentration measurements from air quality monitoring sites and/or spatial data. We gratefully acknowledge the support of the transport company of Zurich (VBZ). The project was financially supported by nano-nera.ch (IRSENS II and Opensense II).

References

Adam, M., Schikowski, T., Carsin, A. E., Cai, Y., Jacquemin, B., Sanchez, M., Vierkotter, A., Marcon, A., Keidel, D., Sugiri, D., Kanani, Z. A., Nadif, R., Siroux, V., Hardy, R., Kuh, D., Rochat, T., Bridevaux, P. O., Eeftens, M., Tsai, M. Y., Villani, S., Phuleria, H. C., Birk, M., Cyrys, J., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M. J., Forsberg, B., de Hoogh, K., Declerq, C., Bono, R., Piccioni, P., Quass, U., Heinrich, J., Jarvis, D., Pin, I., Beelen, R., Hoek, G., Brunekreef, B., Schindler, C., Sunyer, J., Kramer, U., Kauffmann, F., Hansell, A. L., Kunzli, N., and Probst-Hensch, N.: Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis, European Respiratory Journal, 45, 38-50, 2014.

Arfire, A., Marjovi, A., and Martinoli, A.: Model-based rendezvous calibration of mobile sensor networks for monitoring air quality, ~~1-4 Nov. 2015~~IEEE SENSORS 2015, 1-4.

Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P., and Karr, C.: A Cohort Study of Traffic-Related Air Pollution Impacts on Birth Outcomes, Environ Health Perspect, 116, 680-686, 2008.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and regression trees, Chapman & Hall, 1984.

Brunner, J. and Scheller, M.: Luftbilanz Stadt Zurich (Air quality report of the city of Zurich). 2014.

Courtillot, I., Morville, J., Motto-Ros, V., and Romanini, D.: Sub-ppb NO2 detection by optical feedback cavity-enhanced absorption spectroscopy with a blue diode laser, Applied Physics B-Lasers and Optics, 85, 407-412, 2006.

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

EEA: Air quality in Europe - 2016 report. EEA-European Environment Agency, 2016.

Fischer, M., Tuzson, B., Hugi, A., Broennimann, R., Kunz, A., Blaser, S., Rochat, M., Landry, O., Mueller, A., and Emmenegger, L.: Intermittent operation of QC-lasers for mid-IR spectroscopy with low heat dissipation: tuning characteristics and driving electronics, Optics Express, 22, 7014-7027, 2014.

Fuchs, H., Dube, W. P., Lerner, B. M., Wagner, N. L., Williams, E. J., and Brown, S. S.: A Sensitive and Versatile Detector for Atmospheric NO2 and NOx Based on Blue Diode Laser Cavity Ring-Down Spectroscopy, Environ. Sci. Technol., 43, 7831-7836, 2009.

Gehring, U., Gruzieva, O., Agius, R. M., Beelen, R., Custovic, A., Cyrys, J., Eeftens, M., Flexeder, C., Fuertes, E., Heinrich, J., Hoffmann, B., de Jongste, J. C., Kerkhof, M., Klümper, C., Korek, M., Mölter, A., Schultz, E. S., Simpson, A., Sugiri, D., Svartengren, M., von Berg, A., Wijga, A. H., Pershagen, G., and Brunekreef, B.: Air pollution exposure and lung function in children: the ESCAPE project, Environ Health Perspect, 12, 1357-1364, 2013.

Hagemann, R., Corsmeier, U., Kottmeier, C., Rinke, R., Wieser, A., and Vogel, B.: Spatial variability of particle number concentrations and NOx in the Karlsruhe (Germany) area obtained with the mobile laboratory 'AERO-TRAM', Atmospheric Environment, 94, 341-352, 2014.

Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J., and Thiele, L.: Deriving high-resolution urban air pollution maps using mobile sensor nodes, Pervasive and Mobile Computing, 2015. 268-285, 2015.

Jagerska, J., Jouy, P., Hugi, A., Tuzson, B., Looser, H., Mangold, M., Beck, M., Emmenegger, L., and Faist, J.: Dual-wavelength quantum cascade laser for trace gas spectroscopy, Applied Physics Letters, 105, 2014.

Jagerska, J., Jouy, P., Tuzson, B., Looser, H., Mangold, M., Soltic, P., Hugi, A., Broennimann, R., Faist, J., and Emmenegger, L.: Simultaneous measurement of NO and NO2 by dual-wavelength quantum cascade laser spectroscopy, Optics Express, 23, 1512-1522, 2015.

Kebabian, P. L., Herndon, S. C., and Freedman, A.: Detection of nitrogen dioxide by cavity attenuated phase shift spectroscopy, Anal. Chem., 77, 724-728, 2005.

Kehl, P.: GPS based dynamic monitoring of air pollutants in the city of Zurich, Switzerland, 2007. ETH Zurich, 2007.

Mangold, M., Tuzson, B., Hundt, M., Jágerská, J., Looser, H., and Emmenegger, L.: Circular paraboloid reflection cell for laser spectroscopic trace gas analysis, J. Opt. Soc. Am. A, 33, 913-919, 2016.

McManus, J. B., Zahniser, M. S., Nelson, D. D., Shorter, J. H., Herndon, S. C., Jervis, D., Agnese, M., McGovern, R., Yacovitch, T. I., and Roscioli, J. R.: Recent progress in laser-based trace gas instruments: performance and noise analysis, Appl. Phys. B, 119, 203-218, 2015.

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J. J., McLeod, M. W., Hodgson, T. F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J. R., and Jones, R. L.: The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks, Atmospheric Environment, 70, 186-203, 2013.

Mueller, M. D., Hasenfratz, D., Saukh, O., Fierz, M., and Hueglin, C.: Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network, Atmospheric Environment, 126, 171-181, 2016.

Mueller, M. D., Wagner, M., Barmpadimos, I., and Hueglin, C.: Two-week NO2 maps for the City of Zurich, Switzerland, derived by statistical modelling utilizing data from a routine passive diffusion sampler network, Atmospheric Environment, 106, 1 - 10, 2015.

NABEL: Technischer Bericht zum Nationalen Beobachtungsnetz für Luftfremdstoffe (NABEL), 2016.

Osthoff, H. D., Brown, S. S., Ryerson, T. B., Fortin, T. J., Lerner, B. M., Williams, E. J., Pettersson, A., Baynard, T., Dube, W. P., Ciciora, S. J., and Ravishankara, A. R.: Measurement of atmospheric NO2 by pulsed cavity ring-down spectroscopy, J. Geophys. Res.-Atmos., 111, 10, 2006.

OSTLUFT: Die Luftqualitätsüberwachung der Ostschweizer Kantone und des Fürstentums Liechtenstein: Luftqualität 2015, 2016.

Patimisco, P., Scamarcio, G., Tittel, F. K., and Spagnolo, V.: Quartz-Enhanced Photoacoustic Spectroscopy: A Review, Sensors, 14, 6165-6206, 2014.

Pushkarsky, M., Tsekoun, A., Dunayevskiy, I. G., Go, R., and Patel, C. K. N.: Sub-parts-per-billion level detection of NO2 using room-temperature quantum cascade lasers, Proc. Natl. Acad. Sci. U. S. A., 103, 10846-10849, 2006.

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J. M., Gamache, R. R., Harrison, J. J., Hartmann, J. M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R. J., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Mueller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., and Wagner, G.: The HITRAN2012 molecular spectroscopic database, Journal of Quantitative Spectroscopy & Radiative Transfer, 130, 4-50, 2013.

Saukh, O., Hasenfratz, D., and Thiele, L.: Reducing multi-hop calibration errors in large-scale mobile sensor networks, Proceedings of the 14th International Conference on Information Processing in Sensor Networks, Seattle, Washington, 274-285, 2015.

Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevot, A. S. H., and Hueglin, C.: Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, Journal of Geophysical Research: Atmospheres, 112, n/a-n/a, 2007.

Süess, M., Hundt, P., Tuzson, B., Riedi, S., Wolf, J., Peretti, R., Beck, M., Looser, H., Emmenegger, L., and Faist, J.: Dual-Section DFB-QCLs for Multi-Species Trace Gas Analysis, Photonics, 3, 24, 2016.

Therneau, T., Atkinson, B., and Ripley, B.: rpart: Recursive Partitioning and Regression Trees. 2015.

Tuzson, B., Zeyer, K., Steinbacher, M., McManus, J. B., Nelson, D. D., Zahniser, M. S., and Emmenegger, L.: Selective measurements of NO, NO2 and NOy in the free troposphere using quantum cascade laser spectroscopy, Atmospheric Measurement Techniques, 6, 927-936, 2013.

Werle, P., Mucke, R., and Slemr, F.: The limits of signal averaging in atmospheric trace-gas monitoring by tuneable diode-laser absorption-spectroscopy (TDLAS), Applied Physics B-Photophysics and Laser Chemistry, 57, 131-139, 1993.

WHO: Review of evidence on health aspects of air pollution – REVIHAAP project: final technical report. World Health Organization (WHO) Regional Office for Europe, 2013.