Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Changes of NOx in urban air detected with monitoring VIS-NIR field spectrometer during the coronavirus pandemic: A case study in Germany

Paul Naethe, Michael Delaney, Tommaso Julitta

Rhine-Waal University of Applied Sciences, Faculty of Communication and Environment, Friedrich-Heinrich-Allee 25, 47475 Kamp-Lintfort, Germany
JB Hyperspectral Devices, Am Botanischen Garten 33, 40225 Düsseldorf, Germany

HIGHLIGHTS
• A significant drop in urban NO2 emission was reported during the COVID-19 lock-down.
• Simultaneous measurements of down-welling light model NO2 concentration
• Inexpensive field spectrometer system reproduced the drop in NO2 levels.
• Exploited wavebands in visible range model NO2 levels with promising accuracy.
• Monitoring NO2 as risk factor for COVID-19 may also include field spectrometers.

GRAPHICAL ABSTRACT

ABSTRACT

The global outbreak of the coronavirus pandemic has led to a significant reduction of traffic and traffic-related urban air pollution. One important pollutant in this context is NO2. Sudden change in NO2 emissions related to reduction of urban traffic due to infection protection measures can be detected in Düsseldorf, Germany with continuous measurements of down-welling light with a RoX automated field-spectrometer. In comparison to a nearby reference instrument, a waveband around 590 nm was identified as significant for the retrieval in the VIS-NIR spectral range. A decision tree based on principal components which were decomposed from down-welling radiance spectra has been the most robust approach to retrieved NO2 values. Better differentiation of the NO2 value-range is achieved with a partial least square regression model. The results suggest that traffic-related changes of NOx pollution in urban air can be detected through continuous down-welling radiance measurements with inexpensive automated field-spectrometer systems.

1. Introduction

Air quality is critical for the well-being and health of humans living in urban areas. Nitrous oxides (NOx) are produced as anthropogenic pollutant and are found in elevated concentrations in densely populated urban areas. Recent studies suggest that long-term exposure to elevated NO2 has an impact on fatalities due to the coronavirus, especially in areas with low wind speed and low airmass exchange (Coccia, 2020a; Frontera et al., 2020; Ogen, 2020). In particular, stagnation of air pollutants due to the typical climatology of urban areas in the backcountry supports the spread of viral infective agents (Coccia, 2020b). In recent years, efforts have been made to reduce NOx emission of combustion engines as a consequence of the legal obligations and stricter regulations from governments in response to the Volkswagen emission scandal in...
2015. A network of permanent monitoring stations has been established in Germany under state-authority to monitor complete levels of typical air pollutants in urban hotspots. These monitoring efforts provide comprehensive data of NO, NO2 and other typical pollutants at various locations.

As a result of the COVID-19 outbreak in Wuhan, China, and the subsequent global pandemic, most affected countries have responded with varying shut-down measures. In effect, there have been social interactions reduced, economic activities largely halted and personal mobility discouraged. The resulting reduction of anthropogenic activities has also affected the emission of urban air pollutants noticeably (Muhammad et al., 2020; Wang et al., 2020). A better approach to monitoring and regulating urban air pollutants may very well be part of an integral strategy to address the current coronavirus pandemic and reduce environmental risk factors also for future diseases similar to the COVID-19 infection (Coccia, 2020b).

Geostatistical data fusion approaches are promising in combining real-time data from high-density low-cost sensor-networks with high-grade stationary instruments for improved observation of urban air quality (Schneider et al., 2017). Portable quantum cascade laser spectrometers have been shown to accurately record NO2 in ambient air (Hundt et al., 2018). A waveband in the infrared is shown optimal for NO2 retrieval (He et al., 2019). However, previous studies indicate typical absorption cross-sections for opto-electronic detection of the urban air pollutant also in the VIS-NIR range (Zeng et al., 2018). Furthermore, high-grade imaging UV-VIS sensors and UV-VIS point-spectrometers have measured columnar amounts of NO2 in the atmosphere (Park et al., 2019). Thus, it may be possible to consider also inexpensive VIS-NIR spectrometer systems as suitable devices for detecting NO2 pollution in urban air through the continuous monitoring of downwelling light passing through the earth’s atmosphere. Hence, inexpensive VIS-NIR field-spectrometers could support the deployment of widespread monitoring networks of urban air pollution. We observe significant changes in NO2 concentration following the reduction of traffic as an immediate effect of the coronavirus pandemic. This case study explains how continuous measurements of down-welling light with inexpensive field-spectrometers can be exploited to retrieve concentration of urban NOx pollution for the purpose of calibration and validation. The presented analysis showed promising results in that sense with significant correlation and an accuracy of up to 87.3% between retrieved and measured NO2 concentration across three different retrieval models.

2. Data and methods

In the following details are provided on the study area, ancillary data, measures of this study, data analyses and procedures.

- Study area, data and sources

This study was conducted in close proximity to Südring, a 6-lane street in the south of central Düsseldorf (NS1.201745, E6.761245).

Continuous long-term measurements of urban air quality are available as open data from the State Department for Nature, Environment and Consumer Protection North-Rhine Westphalia (LANUV NRW, https://www.lanuv.nrw.de/umwelt/luf/immissionen/berichte-und-trends/einzelwerte-kontinuierlicher-messungen). NO2 levels were extracted from the nearest automated measuring station. Data from the station DDCS, Düsseldorf, Corneliusstraße is available for the time period of interest (Fig. 1).

- Measures of the study

Automated long-term measurements of down-welling light were performed with a RoX monitoring field spectrometer system by JB-Hyperspectral Devices, Düsseldorf, Germany (Fig. 2). The solar-powered instrument was installed for that purpose on a rooftop at the study site.

The data recorded covers the period between December 1st, 2019 and March 29th, 2020. The hyperspectral data was processed from raw digital numbers to calibrated radiances using the FieldSpectroscopyDP (https://github.com/tommasojulitta/FieldSpectroscopyDP) and FieldSpectroscopyPyCC (https://github.com/tommasojulitta/FieldSpectroscopyPyCC) packages in R (R Core Team, 2017).

- Data analysis and procedure

Overall, three statistical models were selected for the retrieval of NO2 concentration levels from hyperspectral radiance data: an unsupervised classification model, a decision tree model and partial least squares regression model. These methodologies are in the named order increasing in complexity and are capable of retrieving better NO2 differentiation from the optical measurements. However, with increasing complexity require those methods more information from the underlying data and thus are more sensitive to propagation of uncertainty from the hyperspectral measurements. The analyses of the hyperspectral data were performed in R (R Core Team, 2017). In particular the packages tidyrverse (Wickham et al., 2019), mdatools (Kucheryavskiy, 2020), rpart (Therneau and Atkinson, 2019) and partykit (Hothorn et al., 2006) were used. Initially, the hyperspectral data was aggregated by a one-hour interval and the mean of each waveband calculated for each interval step to match the hourly NO2 measurements.

A binary class was introduced, labeled before COVID-19 (bC) and after COVID-19 (aC). A noticeable drop in the NO2 time-series was identified as breakpoint for the separation of this binary class (Fig. 1). This breakpoint has been validated as one day after the TV-oration of German Chancellor Angela Merkel on the evening of March 18th, 2020, in which she announced contact prohibition and advanced protective measures in response to the coronavirus pandemic. Both, the hyperspectral radiance time series and NO2 time-series were combined in a large dataset. As suggested in Grolemund and Wickham (2016), the data was transformed in an effort to improve accessibility. A spectral window between 550 nm and 680 nm (Fig. 2) was selected for further analysis in agreement with the spectral range of the known NOx absorption cross-section (Zheng et al., 2018) and to excluded unrelated spectral regions as potential source of error. An unsupervised classification model was built with a conditional inference tree, using the hyperspectral radiance data to predict the binary class response. The model performance was recorded and a confusion matrix created. The positive predictive values of the model were reported separately for each group and the entire set.

Furthermore, the down-welling radiance data was decomposed into principal components (PCs). These principal components served as the
predicting variables for an unsupervised decision tree model for the regression with the measured NO2 values. A penalty was introduced to limit tree complexity and avoid over-fitting. With $R^2$ increasing by at least 0.01 per additional level can the model discriminated 10 discrete NO2 levels. Tree structure and most significant variable are reported in Fig. 3. The model was built of nine PCs at nine nodes, with PC2 as most significant for the differentiation of NO2 which is reflected by its position at the top of the decision tree. The decision tree model has been used to predict NO2 levels from the actual measurements.

The predicted NO2 values were compared with the true measured values and confidence of determination ($R^2$) was calculated for a linear regression. In addition, intercept and slope for the linear fit and the p-value for significance of correlation are reported. Furthermore, the F-test for ratio of variance and its p-value for significance are reported. Moreover, partial least square regression was employed with the down-welling radiance data to predict actual NO2 values with a supervised model. This model is built on the assumption of a linear regression between NO2 concentration and a covariance in the mixture of signals in the down-welling light measurements. Based on this correlation between NO2 levels and down-welling light measurements was the variable space of the calibration matrix decomposed into optimized latent variables (components) in an iterative process. The resulting components of the spectral radiance data were then utilized in a model to predict the actual NO2 levels from the down-welling light measurements with the instruments. The full dataset was split even into training and testing data for model training and cross-validation. The Root Means Square Error (RMSE) was calculated as following:

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (\hat{x}_i - x_i)^2}$$

with $\hat{x}_i$ being the actual measured NO2 value and $x_i$ being the predicted NO2 value was reported as in Eq. (1) for increasing number of components, and thereby an increasing model complexity. For prediction, the number of components with lowest RMSE in the cross-validation was selected. The absolute regression coefficients were reported for this model complexity with respect to wavelength. Finally, the predicted NO2 values were compared with the actual measurements across the entire dataset through linear regression, reporting $R^2$ and RMSE of prediction.

### 3. Results and discussion

Hyperspectral down-welling radiance time-series were exploited to retrieve NO2 from urban sources. The results from three different retrieval approaches were compared to actual measurements of the urban monitoring. Results are presented according to the previously described methodology. The inter-quartile ranges (IQR) of NO2 values of the group before COVID-19 (bC) is above the one of group after COVID-19 (aC), both IQR do not overlap. Both groups, bC and aC are therefore significantly different (see Fig. A1 in Appendix A). On the weekend after the announcement of German Chancellor Merkel, on March 21st and 22nd, were the lowest daily maximum values recognized in the entire NO2 series (Fig. 1). These findings suggest reduced NO2 emissions in response to the official recommendation for people to stay home.
Further, a conditional inference tree based on hyper spectral down-welling radiances suggested a waveband around 590.56 nm as most significant for the classification of bC and aC within a 99.9% confidence interval. The absorption cross-section of both NO2 and NO3 presented in Zheng et al. (2018) overlap with this waveband around 590 nm. Hence, cooccurrence of NO2 and NO3 is assumed on the basis of traffic-related emission sources.

The model classified the NO2 correctly from down-welling light in 91.9% of the bC cases and in 73.3% of the aC cases. Overall, 87.3% of the cases are distinguished correctly, based on hyperspectral down-welling radiance data (Table 1).

The decision tree model has been employed to correlate NO2 values with combinations of PCs, derived from continuous downwelling light measurements. Fig. 4 illustrates the regression between predicted and actual measured NO2 values.

The F-test indicates a significant difference within a 99.9% confidence interval between the decision tree model and an intercept-only model for the prediction of NO2 values (Table 2). Therefore, it can be concluded that the specified decision tree model describes the measured data better than a linear model with no predictors. Hence, the model can reconstruct absolute NO2 values with $R^2 = 0.67$ from hyperspectral data, showing a significant linear correlation within a 99.9% confidence interval (Table 2). Slope and intercept indicate furthermore a one-to-one relationship between measured and predicted NO2.

Table 1
Confusion matrix for prediction of nominal NO2 levels after COVID-19 (aC) and before COVID-19 (bC) with the conditional inference tree. The percentage of true positive classification is given per class and per total.

|       | aC | bC | Positive predictive value |
|-------|----|----|---------------------------|
| aC    | 22 | 8  | 73.3%                     |
| bC    | 8  | 88 | 91.6%                     |
| total | 30 | 96 | 87.3%                     |

Partial least squares regression was applied, relating NO2 to downwelling radiance measurements, and the performance reported in Fig. 5. During the model calibration decreases the error of prediction with increasing model complexity, reaching an almost perfect fit of the training data. However, this is a typical case of overfitting as the error of prediction increases rapidly with increasing model complexity during the cross-validation. The optimal number of components for prediction with minimal error in the cross-validation process is at 10 components. The regression coefficients are plotted in absolute values to improve readability. These coefficients can also be negative as their sign arbitrarily depends on the orthogonal loadings. The highest regression coefficient is computed for a waveband around 590 nm and deemed the most influential for the prediction of NO2. The 10 components model predicts measured NO2 with $R^2 = 0.45$ and RMSE = 14.22 μg/m$^3$ with respect to the actual measured values.

Best accuracy of the three statistical models tested for the retrieval of NO2 was achieved with the classification through a conditional inference tree. The presented classification approach, on the other hand, is limited due to the repeated temporal occurrence of similar values within the outliers in each group. This overlap is especially noticeable for the group aC, while the group bC can be distinguished through a unique value-range. As a consequence, misclassification of spectra that would actually fall into aC is more likely. This, in turn, is directly reflected in the ambiguity of the conditional classification. The positive predictive value for bC is significantly higher in comparison to aC. To reduce classification bias by cooccurrence with other trends in the hyperspectral data the classification should be carried out with an extended time-series, comparing similar episodes in subsequent years. Unfortunately, due to the limited timeframe of field spectrometer data available in this instance, this has not been possible. For the complexity of changes in the atmospheric transfer across the entire spectral range statistical de-trending methods were considered unsuitable. The unsupervised regression model presented remains unaffected by those limitation. In contrast to comparing classes split at similar NO2 values. The presented unsupervised regression model shows the highest $R^2$ of all three investigated approaches but remains rather crude in value-resolution as only 10 levels of NO2 values can be effectively differentiated from hyperspectral PCs.

In comparison, the presented supervised regression is more capable of reproducing the actual value-range in NO2 but shows a lower $R^2$ value. A cross validation process in the supervised approach is testing the model’s robustness with previously unknown data from repeated random segments and minimizes over-fitting. However, the algorithm
is highly sensitive to noise in the hyperspectral radiance data which is also reflected in the model’s spectral regression coefficients. Moreover, a waveband around 590 nm was recognized as the most influential in the supervised model. This outcome aligns with the previous findings in the unsupervised classification and is in accordance with the absorption cross-section for NOx in the VIS-NIR spectral range. In Fig. 2 a small absorption feature is noticed around 590 nm in the down-welling light spectrum. However, the immediate exploitation remains tricky as there are multiple atmospheric influences in this spectral region.

Nevertheless, the observed significant drop in urban air pollution with NOx in this case study can be considered a beneficial byproduct of the lockdown and reduce environmental risk factors for COVID-19 and similar diseases, threatening to human health (Fattorini and Regoli, 2020). Our results encourage further investigation of field-spectrometers as inexpensive instruments towards a wide-spread monitoring network of anthropogenic air pollution as a part of an integrated strategy to reduce further environmental risk factors for the coronavirus and similar pandemics in the future.

4. Conclusion

Three different classification and regression approaches were tested in this case study to retrieve information on urban air pollutants from continuous down-welling radiance measurements with an automated monitoring field-spectrometer system during the coronavirus pandemic. Significant differences in urban NO2 emissions were reported before and after severe reduction of urban traffic due to measures in response to the coronavirus pandemic which were enacted by governmental authorities in Germany. The investigated methods reproduce these differences in measured NO2 successfully with relatively inexpensive equipment. As an alternative, physical modelling based on mechanistic radiative transfer models can be considered for the retrieval of urban air pollutants from down-welling hyperspectral measurements. However, further direct exploitation of absorption features of NO2 is considered very tricky, as they are typically overlaid with other absorption features of atmospheric trace gases and aerosols. In addition, geographical position is known to influence the stagnation of air pollutants under certain conditions which has to be considered as a limitation to this investigation. It has to be further noted that the location of measurements for NO2 and down-welling light were not identical due to the availability of data. Therefore, uncertainties with respect to the spatial distribution have to be considered. Further limitation from windspeed as a variable for airmass exchange and the distribution of NO2 in different air layers need to be considered. Additional experiments for calibration and validation over an extended period of time with a dedicated instrumental setup of a monitoring field spectrometer system in close proximity to conventional measurements of urban air pollutants are encouraged. Future efforts should also consider the windspeed, wind direction and the resulting dispersion of NO2 in the air as influencing variables. Nevertheless, in this context we consider the potential of inexpensive, automated field-spectrometers a valuable contribution towards wide-spread monitoring of urban air pollutants, especially since these systems are easy to handle, cheap to maintain and observe a large spectral range containing various other information.

CRediT authorship contribution statement

Paul Naethe: Conceptualization, Methodology, Data curation, Software, Formal analysis, Writing - original draft, Visualization. Michael Delaney: Software, Validation, Writing - review & editing. Tommaso Julitta: Investigation, Validation, Funding acquisition, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 2

| Statistic                  | Test value |
|----------------------------|------------|
| F-test                    | 253.3      |
| F-test p-value            | <2 × 10^{-16} |
| Intercept of linear regression | 0         |
| Slope of linear regression | 1         |
| R^2                       | 0.67       |
| Correlation p-value       | <2.2 × 10^{-16} |

Fig. 5. Root Mean Square Error (RMSE) of calibration (blue) and cross-validation (red) with respect to number of components used – (a). Absolute regression coefficients of the 10 components model with respect to wavelength – (b). Benchmarking prediction of NO2 values with the 10-components model – (c).
Acknowledgement

This work was supported in part through the joint research project “mDRONES4rivers” by the Federal Ministry of Transport and Digital Infrastructure – BMVI (19F2054A-D). Special thanks to Andreas Burkart for inspirational discussions and contributing excellently designed monitoring field-spectrometer systems.

Appendix A

Fig. A1. Panel (a) shows boxplot with two groups of NO2 values, before COVID-19 (bC) in green and after COVID-19 (aC) in red. Panel (b) shows the conditional inference tree for the classification of nominal NO2 levels from hyperspectral down-welling measurements.

References

Coccia, M., 2020a. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Sci. Total Environ. 729, 138474. https://doi.org/10.1016/j.scitotenv.2020.138474.

Coccia, M., 2020b. How high wind speed can reduce negative effects of confirmed cases and total deaths of COVID-19 infection in society. SSRN Electron. J. https://doi.org/10.2139/ssrn.3603380.

Fattorini, D., Regoli, F., 2020. Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy. Environ. Pollut. https://doi.org/10.1016/j.envpol.2020.114732.

Frontera, A., Cianfanelli, L., Vlachos, K., Landoni, G., Crennora, G., 2020. Severe air pollution links to higher mortality in COVID-19 patients: the “double-hit” hypothesis. J. Inf. Secur. https://doi.org/10.1016/j.jinf.2020.05.031.

Grolemund, G., Wickham, H., 2016. R for Data Science. WWW Document. O’Reilly Media URL. http://r4ds.had.co.nz/. (Accessed 10 February 2018).

He, X., Xu, X., Zheng, Z., 2019. Optimal band analysis of a space-based multispectral sensor for Urban Air Pollutant Detection. Atmosphere (Basel) https://doi.org/10.3390/amt10100631.

Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Stat. 15, 651–674. https://doi.org/10.1198/106186006X133933.

Hundt, P.M., Müller, M., Mangold, M., Tuzson, B., Scheidegger, P., Koepsler, H., Higlin, C., Emmenegger, L., 2018. Mid-IR spectrometer for mobile, real-time urban NO2 measurements. Atmos. Meas. Tech 11, 2669–2681. https://doi.org/10.5194/amt-11-2669-2018.

Kucheryavskiy, S., 2020. mdatools – R package for chemometrics. Chemom. Intell. Lab. Syst. 198. https://doi.org/10.1016/j.chemolab.2020.103937.

Muhammad, S., Long, X., Salinan, M., 2020. COVID-19 pandemic and environmental pollution: a blessing in disguise? Sci. Total Environ. 728, 138820. https://doi.org/10.1016/j.scitotenv.2020.138820.

Ogen, Y., 2020. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. Sci. Total Environ. 726, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605.

Park, H.-J., Park, J.-S., Kim, S.-W., Chong, H., Lee, H., Kim, H., Ahn, J.-Y., Kim, D.-G., Kim, J., Park, S.S., 2019. Retrieval of NO2 column amounts from ground-based hyperspectral imaging sensor measurements. Remote Sens. 11, 3005. https://doi.org/10.3390/rs11243005.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Therneau, T.M., Atkinson, E.J., 2019. rpart: recursive partitioning and regression trees. [WWW document]. R Packag. version 4.1-13. URL. https://cran.r-project.org/package=rpart. (Accessed 28 May 2020).

Wang, Y., Yuan, Y., Wang, Q., Liu, C.G., Zhi, Q., Cao, J., 2020. Changes in air quality related to the control of coronavirus in China: implications for traffic and industrial emissions. Sci. Total Environ. 731, 139133. https://doi.org/10.1016/j.scitotenv.2020.139133.

Zheng, K., Zheng, C., Zhang, Y., Tittel, F.K., 2018. Review of incoherent broadband cavity-enhanced absorption spectroscopy (IBBCEAS) for gas sensing. Sensors (Basel) https://doi.org/10.3390/s18113646.