Evaluation and Field Calibration of a Low-Cost Ozone Monitor at a Regulatory Urban Monitoring Station

Mauro Masiol,[1] Stefania Squizzato,[1] David Chalupa,[2] David Q. Rich,[1,2] Philip K. Hopke,[1,3, †]

[1] *Department of Public Health Sciences, University of Rochester Medical Center, Rochester, NY 14642*
[2] *Department of Environmental Medicine, University of Rochester Medical Center, Rochester, NY 14642, United States*
[3] *Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY 13699*

ABSTRACT

The performance of a low cost ozone monitor (Aeroqual Series 500 portable gas monitors coupled with a metal oxide sensor for ozone; model OZL) was assessed under field conditions. Ten ozone monitors were initially calibrated in clean-air laboratory conditions and tested at controlled ozone concentrations of 5 to 100 ppb. Results showed good linearity and fast response with respect to a conventional research-grade ozone monitor. One monitor was then co-located at a regulatory air quality monitoring station that uses a U.S. federal equivalent method (FEM) ozone analyzer. Raw data from the Aeroqual monitor collected over 4 months (June-October) at a 10-minute time-resolution, showed good agreement (r^2=0.83) with the FEM values but with an overestimation of ~12%. Data were averaged to different time resolutions; 1 h time averaged concentrations showed the best fit with the FEM results (r^2=0.87). An analysis of the ratio of FEM/monitor concentrations against chemical and meteorological variables suggested the potential of interferences due to temperature, relative humidity, nitrogen oxides, and volatile organic compounds. Three correction models using temperature, humidity, and nitrogen dioxide (NO_2) were then tested to better relate the monitor concentrations to the FEM values. Temperature and humidity are two variables commonly available (or easily measurable) at sampling sites. The model (#3) that added NO_2 did not provide a substantial improvement in the fit. Thus, the proposed models with only temperature and humidity can be easily adopted and adapted by any user. The corrected data explained up to 91% of the variance and showed statistically significant improvement of the goodness of fits as well as decreased influence of the interfering variables on the diurnal and weekly patterns. The correction models were also able to lower the effect of seasonal temperature changes, allowing the use of the monitors over long-term sampling campaigns. This study demonstrated that the Aeroqual ozone monitors can return "FEM-like" concentrations after appropriate corrections. Therefore, data provided by a network of monitors could determine the intra-urban spatial variations in ozone concentrations. These results suggest that these monitors could provide more accurate human exposure assessments and thereby reduce exposure misclassification and its resulting bias in epidemiological studies.

Keywords: Semiconductor gas sensor, tropospheric ozone, urban air pollution, air pollution exposure

[†] Author to whom correspondences should be addressed. Email: phopke@clarkson.edu

1. INTRODUCTION

Tropospheric ozone (O_3) is a greenhouse gas and an air pollutant (Stevenson *et al*., 2013; Cooper *et al*., 2014; Monks *et al*., 2015) known to be harmful to human health (Jerrett *et al*., 2009; Bell *et al*., 2014; Turner *et al*., 2016) and ecosystems (Fowler *et al*., 2009; Ainsworth *et al*., 2012). In U.S., the National Ambient Air Quality Standards (NAAQS) set the limit values for the protection of public health (primary standard) and public welfare (secondary standard) of six principal "criteria" air pollutants, including ozone. Concentrations of ozone are measured using federal "reference" or "equivalent" methods (FRM and FEM, respectively) in accordance with Code of Federal Regulations (40 CFR Part 53; USEPA, 2017). Compliance with NAAQS within major cities is routinely evaluated at one or a few static urban stations. The spatial coverage of monitoring networks is therefore insufficient to capture the spatial variability due to the effect of major roadways, complex terrain, urban heat island effects, and the locations and strengths of local sources of ozone precursors (i.e. nitrogen oxides (NO_x=NO+NO_2), carbon monoxide (CO), and reactive volatile organic compounds (VOCs) of biogenic and anthropogenic origin.

Recent advances in micro-scale technology have made available inexpensive and reliable sensors and low power electronic circuits and memory, allowing the development of a series of low-cost (and relatively low-cost) air quality monitors. These devices are much less expensive than research-grade instruments, have low power requirements, and are physically smaller and lighter (mostly portable). Although the use of this technology increases the potential spatial resolution of monitoring networks, current monitors are not designed to meet rigid performance standards and generate less accurate data than research-grade instruments (White *et al*., 2012; Snyder *et al*., 2013; Kumar *et al*., 2015). This limitation can be addressed by careful calibration of the units and post-processing of the raw data before they can be used for scientific purposes.

A cost-efficient ozone monitor using gas-sensitive semiconducting oxide (GSS) technology was co-located over 4 months to a routine air quality monitoring station in a medium-sized city of the Northeast U.S. (Rochester, NY). Ozone concentrations were compared with those measured by a FEM instrument and an extensive series of additional air pollutants and weather data measured with scientific-grade or conventional instruments. A series of chemometric approaches were adopted to (i) calibrate the monitor, (ii) evaluate the effects and interferences with other measured air pollutant species under field conditions, (iii) correct the data for interferences to return unbiased concentrations as similar as possible to the reference

method (FEM), and (iv) discuss strengths and limits of using GSS technology for scientific purposes.

2. MATERIALS AND METHODS

2.1 Study area

Rochester, NY, (~210,000 inhabitants, 2010 Census) is the center of the Greater Rochester metropolitan area (~1.1 million inhabitants) and lies on the southern shore of Lake Ontario. It is typical of NE U.S. moderate sized, urban areas. Road traffic is a major source of CO, NO_x, and $PM_{2.5}$ (Figure S1 in the supplemental information file). This study was performed at the regulatory air quality station in Rochester (ROC; USEPA 36-055-1007) operated by the New York State Department of Environmental Conservation (NYSDEC). The site lies ~300 m from the intersection of two major highways with an average traffic of ~ 230,000 vehicles/day (Figure S1). Diesel-powered trains operate on a mainline track affect the site with a switching yard about 1.5 km NW of the site. Regional advection of polluted air masses from Buffalo (NY), the Ohio River Valley, eastern coast of the U.S. and occasionally from Toronto (ON) also affect local air quality (Emami *et al*., 2018).

2.2 Experimental

Ten Aeroqual (Auckland, New Zealand) Series 500 portable gas monitors coupled with metal oxide (WO_3) GSS sensors for ozone (model OZL) were purchased in April 2017. Technical details of these monitors are discussed elsewhere (Aliwell *et al*., 2001; Williams *et al*., 2009; 2013) and summarized in Table S1. The sensor operates in the 0 to 0.5 ppm O_3 concentration range and has a minimum detection limit (MDL) of 0.001 ppm with an accuracy of 0.008 ppm over the 0 to 0.01 ppm range and <±10% for the rest of the range. Preliminary tests of this type of sensor under field conditions reported promising results. Bart *et al* (2014) reported that differences (GSS-reference instrument) in hourly average ozone concentrations were normally distributed with a mean = -0.001 ppm and standard deviation = 0.006 ppm. Lin *et al*. (2015) reported a high coefficient of determination (r^2=0.91) with values measured with a reference ultraviolet absorption O_3 analyzer.

In this study, one monitor was placed on the roof of the NYSDEC ROC station (Figure S2) in a waterproof plastic-fiberglass enclosure with two 90° bend inlets (2 cm in diameter) and a

USB-powered fan (4500 RPM) to promote air throughput (Figure S2). A single layer of screening was placed on each inlet to prevent coarse debris and insects from entering the enclosure. A low-cost PM monitor (Speck, Airviz Inc., PA) that also has a temperature sensor was installed in a similar box side-by-side to the ozone monitor (Zikova *et al*., 2017). This latter monitor was used to measure the air temperature inside the enclosures, i.e. monitoring the potential diurnal temperature from increasing solar irradiance heating. The monitor was powered with a 12 V DC power supply, but also had a Li-ion 2700 mA h^{-1} battery to eliminate the effects of short-term power outages (up to ~8 h). The sampling campaign extended over 4 months (June 9 to October 11, 2017). Date were collected with a time resolution of 10 minutes. Periodic checks of the fan operation, downloads of the data, and cleaning of the inlets were performed throughout the sampling campaign.

2.3 Air quality and weather variables

Concentrations of CO, NOx, total reactive nitrogen (NO_y), SO_2, O_3, PM_{10} and $PM_{2.5}$ were routinely measured by NYS DEC in accordance with federally mandated methods at a 1-minute time resolution. NO_2 was estimated as NOx-NO. $PM_{2.5}$-bound equivalent black carbon (BC) was measured using an aethalometer (Magee model AE-22). Particle number concentrations (PNC) were measured with a scanning mobility particle spectrometer (SMPS) at 5 min resolution time. The number concentrations were split into three ranges roughly representative of nucleation (NUC: 11-50 nm), Aitken nuclei (AIT: 50-100 nm) and accumulation (ACC: 100-470 nm) particles. Details are reported in Table S2. FEM ozone was detected with a Teledyne API T400 photometric analyzer based on ultraviolet absorption (automated equivalent method EQOA-0992-087) (USEPA, 2017; NYSDEC, 2017). The FEM analyzer is regularly calibrated weekly to a secondary transfer standard reference, a photometric O_3 calibrator/analyzer (Teledyne, Advanced Pollution Instrumentation, model API 703E). This secondary reference standard is regularly checked (bi-annually) with a NIST primary reference O_3 standard in Albany, NY, as per the Ozone Transfer Standard Guidance Document 10/2013 (USEPA, 2013).

Meteorological data (air temperature (°C), relative humidity (RH, %), barometric pressure (hPa), wind speed (m s^{-1}) and direction) were measured at a 1-h time resolution. Since ROC wind data are potentially affected by surrounding building and street canyon effects, the same set of weather variables (including precipitation) measured at the Greater Rochester

International Airport (KROC) at 1-h intervals were retrieved from the NOAA National Climatic Data Center. Relative humidity values were estimated using data provided by the dense network of personal weather stations across the Monroe County. The reliability of data provided by these personal stations was carefully evaluated and checked with the help of the modeled humidity data provided by the NOAA-NCEP North American Regional Reanalysis (NARR) model at 32 km spatial resolution (every 3 hours).

Additional 3-hour time-resolved meteorological variables estimated by meteorological models were retrieved from the NOAA' National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR; Mesinger *et al*., 2006). These results included downward long- and short-wave radiation fluxes (W m^{-2}), upward long- and short-wave radiation fluxes (W m^{-2}), planetary boundary layer height (m), and forecast total cloud cover (%). The NARR spatial resolution is ~32 km. Final data were calculated as averages within a circular buffer of radius 32 km from the sampling station coordinates.

2.4 Data analyses

Air pollutant data were matched with wind data to investigate the potential local source locations by using polar-plots and polar-annuli. Briefly, polar-plots present the statistics of variables by mapping wind speed and direction as a continuous surface with the surfaces calculated using smoothing techniques (Carslaw *et al*., 2006). Polar-annuli allow mapping wind direction and time of day as continuous surfaces and can present conditional bivariate probability function (CBPF; Uria-Tellaetxe and Carslaw, 2014) or Pearson' correlations between pair of variables (Grange *et al*., 2016). The linear trends of the FEM/Aeroqual ratios throughout the sampling campaign were assessed with the Theil-Sen nonparametric estimator of slope (Theil, 1950; Sen, 1968) that assumes monotonic linear trends and is robust against outliers. Slopes were computed over daily-averaged data (calculated when at least 75% of data are available).

3. RESULTS AND DISCUSSION

3.1 Initial laboratory calibration

Although the sensors used in this study were new with factory calibration certificates, the 10 monitors were calibrated under laboratory conditions prior to the co-location campaign. Historical data measured in Rochester (Emami *et al*., 2018; Squizzato *et al*., submitted) indicate

that summertime hourly ozone peaks never exceeded 100 ppb after 2012. Consequently, the span was calibrated at 100 ppb through a 24 h co-location with an ultraviolet photometric ozone analyzer (Model 49i, Thermo Scientific, Franklin, MA; automated equivalent method EQOA-0880-047). This laboratory reference instrument was calibrated using the same secondary API 703 used to calibrate the FEM analyzer at ROC. Ozone was generated in a clean-air room (trace NO and NO_2; and PNC <100 particles/cm^3) with a Corona spark discharge (model V5-0, Ozone Research and Equipment Corp., Phoenix, AZ) coupled with an ozone calibrator (model 1008-PC, Dasibi Environmental Corp., Glendale, CA). After the span calibration, the monitors were operated overnight with laboratory clean-air O_3 concentrations (<5 ppb). They were checked again the following day at 50 and 100 ppb. Very good agreement was found (mean ± std. deviation of reference instrument/Aeroqual ratio at 100 ppb= 0.99 ± 0.02). The linearity of the Aeroqual monitor responses was again tested at the end of the campaign with the same laboratory procedure for >2 hours and 3 concentration steps (5, 50 and 100 ppb). Results indicated very good linear response to the varying ozone concentrations (r^2=0.99 using 132 readings). However, there were small shifts in the slopes of the response curves from the initial calibration.

3.2 Field co-location

One monitor was co-located at the ROC site with the FEM monitor. Figures S3 and S4 show the boxplots and diurnal profiles of variables measured at ROC. The ambient air temperature recorded during the co-location study (average 20.8°C, range 5-35°C) was within the operating range of GSS sensors (0-40°C) as well as the temperature inside the enclosure (average 21.9°C, range from 0 to 38°C). However, ~19% of data were collected under RH levels exceeding the operating range (>90%), mostly occurring overnight (1-5 am; Figure S5).

During the co-location study, the average ozone concentrations measured with FEM and Aeroqual were 31 and 34 ppb, respectively. The average FEM/Aeroqual ratio was 0.88. Since the monitor was initially calibrated under clean air conditions, the ~12% overestimation is an indication of the interfering substances that affect the GSS sensor sensitivity under field conditions. Aeroqual overestimation (lower ratios) was more frequent during overnight periods (Figure S4), particularly in the early morning (5-7 am local time), when lowest air temperature

and highest RH were observed. A fast increase of FEM/Aeroqual ratio occurred during 7-10 am, concurrent to the morning rush hours for road traffic (peaks of nitrogen oxides, CO and BC; Figure S4), and increasing solar irradiance.

Pearson product-moment correlation coefficients were calculated among the available variables to explore the linear relationships at different time resolutions. Results are arranged as correlograms in Figure S6. The FEM/Aeroqual ratio is moderately ($r>0.6$) positively correlated with FEM ozone concentration, temperature (both ambient air and enclosure) and weakly to moderately ($0.4<|r|<0.6$) negatively correlated with RH at both 10 min and 1 h time-resolutions. The positive correlation with temperature reflects the known decrease in the GSS sensors sensitivity with increased temperature. Since the sensitivity of the monitor was highly linear under lab conditions ($R^2=0.99$ over the 5-100 ppb concentration range), the correlation of the ratio with the FEM ozone concentrations likely reflects the correlations between O_3 and air temperature ($r= 0.73$ and 0.74, for 10-min and 1-h average data, see Figure S6) rather than a non-linear response of GSS sensors. The potential interference of other atmospheric oxidants (possibly radicals) cannot be disregarded. The moderate correlation with RH, but not absolute humidity, is another indication of the temperature effects on the sensor sensitivity. The sensitivity of GSS to ambient water vapor was previously reported by Bart et al. (2014). However, this study found a small systematic error associated with variation of atmospheric relative humidity.

GSS sensors are sensitive to several gaseous species, including NO_2 and VOCs. During the time of this study, nitrogen oxides were largely emitted by road traffic (Figure S3) along with anthropogenic VOCs (with the additional emissions from building and domestic heating in the colder period at the end of the study period). In addition, highly reactive biogenic VOCs (isoprene and terpenes) are present, particularly during the summer (Palmer *et al*., 2003; Sindelarova *et al*., 2014). While NO_2 leads to positive artifacts, VOCs (e.g., butane, heptane, propane, toluene, and particularly alkenes) may cause a reduction in the GSS sensitivity (Aeroqual, 2018). The VOCs are reactants that participate in ozone formation and react with oxidants in many atmospheric processes (Monks, 2005; Seinfeld and Pandis, 2016; Monks *et al*., 2015). Unfortunately, VOCs are not measured hourly at the station. Therefore, their effects on the sensor cannot be evaluated directly. Since Lin *et al*. (2015) reported a limited effect of NO_2 on GSS sensors, the poor ($0.3<|r|<0.4$) negative correlations with NO-NO_2-NO_y likely reflect the

simultaneous effects of multiple species (mostly VOCs) rather than the sole interference of nitrogen oxides.

The Theil-Sen nonparametric estimator of slope was computed on the daily-averaged FEM/Aeroqual ratios to detect the presence of significant slopes in the data that may be indicative of potential instrumental sensitivity drift. Figure S7a shows a statistically significant (p<0.001) negative slope (-0.016/week; 95% confidence interval: -0.021 to -0.011) indicating an increase in the GSS sensitivity throughout the sampling campaign. This result may be biased by air temperatures variations from June to October that showed a similar negative trend.

3.3 Calibration using FEM data

The linear response of the GSS sensor to ozone concentrations over the 5 to 100 ppb range, assessed under laboratory conditions indicated that a simple linear regression between concentrations measured by the monitor (dependent variable) and the FEM instrument (independent variable) is likely to be sufficient for calibration. Figure 1a-e shows results from the linear regression models computed from 10 min time resolution data as well as from increasingly longer time averaged data (from 20 minutes to 1 day). Figure 1f presents the changes in the coefficient of determination and regression coefficients at the various time resolutions and the relative cross-correlations over 24 h lags. The results indicate that the Aeroqual and FEM have a similar response rate to ozone concentrations (cross-correlations peaked at 0 lags using 10 min time intervals and over all the time resolutions). The Aeroqual data were able to explain between 74% and 87% of variability of FEM ozone, with a maximum r^2 for the 1-h interval data. Regression slopes are also higher for 30 min to 2 h average data and have smaller intercepts.

Since the 1-h data provided the best fit with the FEM data and this time resolution is commonly adopted by routine regulatory monitoring networks, the calibration of Aeroqual was calculated for the 1-h time-averaged data. This procedure returned FEM-calibrated data, i.e. Aeroqual data with null intercept, unitary slope and explaining 87% of the original variance (r^2=0.87). A better estimation of the average model prediction errors was assessed by the root-mean-square error (RMSE) and mean absolute error (MAE), two measures of the average differences (in ppb) between values measured by FEM and predicted by the linear model. Results show an average prediction error of 4 ppb (MAE) and 5 ppb (RMSE). The calibration

was able to approximate the average diel and weekly patterns in the FEM data (Figure 2). However, the calibrated data still had a general underestimation during early mornings and on Sundays and an overestimation during the afternoon and on week days. The Theil-Sen slope analysis (Figure S7b) found a lower negative slope of the FEM/Aeroqual ratios (-0.011/week; 95% confidence interval: -0.02 to 0.002). However, the trend was still statistically significant at $p<0.05$.

3.4 Presence of an edge

Figure 1c also shows the presence of an edge splitting record with unusually high FEM/Aeroqual ratios (n=64, 2.3% of total collected data with 1-h time resolution). Figure S8a highlights the edge and displays the occurrence of these events by day and by hour of day. Most of the "high-ratio" events (HRE) occurred at the beginning of the sampling campaign (middle of June), a few occurrences (4 records) on July 6th. No events were detected between August and October (Figure S8b). All HRE occurred during daytime, peaking around noon (Figure S8c). Figure S8 also shows that the regression r^2 increases from 0.87 to 0.9 by excluding these HRE observations.

Figure S9 shows the differences between the "non-edge" data and HRE for selected variables with Kruskal–Wallis one-way analysis of variance tests. Although these tests indicated statistically significant differences ($p<0.05$) between the HRE and normal measurements, these differences were not able to fully explain the presence of the edge, i.e. HRE variability is within the variance of the "non-edge" observations. Wind directionality analysis was performed using polar-plots (Figure S10) and polar-annuli (Figure S11) to investigate possible interferences from local sources or peculiar wind regimes causing the HREs. The highest averages and the highest CBPF probabilities for the FEM/Aeroqual ratio were observed for high winds (speed >6 m/s) blowing from SW during the daytime with similar results for the FEM O_3 concentrations. No unusual patterns were observed with respect to the correlations of FEM/Aeroqual ratio with FEM O_3, air temperature, RH, or downward shortwave irradiation flux (Figure S12). However, careful examination of the HRE occurrences in the bivariate plot of enclosure temperature vs. RH (Figure S13) clearly shows that the HREs occurred when there were high temperatures and low RH. The reason(s) for this combination driving these unusual values is not understood.

3.5 Corrections

The calibration with FEM ozone through linear regression was sufficient to achieve a reasonable goodness of fit. However, the presence of the HREs potentially affected by temperature and humidity (Figure S13) could be accounted for to better approximate Aeroqual data to FEM values. A series of subsequent approaches were tested to correct the data for temperature and RH. The choice of the best method depends on several factors: (i) atmospheric species having effects on the GSS sensitivity are not always measured where the monitors are deployed; (ii) the interplay of interfering species may cause different artifacts (positive, negative, or mixed) that are difficult to separate and adjust for; (iii) the effect of the interfering effects may not be linear; and (iv) there can be multicollinearity between independent variables that may obscure the statistical significance of the models.

The first model (model 1) developed linear relationships of the Aeroqual O_3 with FEM O_3, enclosure temperature, and RH. The results are reported in Table 1 and Figure 2b. They show an overall improvement in the goodness of fit from the initial calibration results. The adjusted r^2 improved to 0.89, the partial F-test indicated statistically significant improvement of the fit at $p<0.001$, RMSE and MAE decreased, and all the independent variables were highly statistically significant ($p<0.001$) in the model. Model 1 also provided a better fit of the daily averages (RMSE=0.4) but not the weekly pattern and resulted in a further decrease in the Theil-Sen slope (Figure S7).

Figure S14 shows the relationships of selected variables against the differences between the concentrations measured by the Aeroqual monitor and the FEM, suggesting that the relationship with temperatures (both ambient air and enclosure) is not linear. Thus, a second model (model 2) was built using the enclosure temperature as a polynomial function. The best results were found by using a polynomial of second order (Table 1). Model diagnostics indicated a further improvement with respect to model 1 (adjusted $r^2=0.90$, RMSE=4.4 ppb and partial F-test $p<0.001$). Model 2 also exhibited an improvement of the fit of weekly-averaged data (Figure 2) and a non-statistically significant Theil-Sen slope ($p>0.1$). This latter result confirms that model 2 successfully corrected the negative effects of temperature on the GSS sensors. Therefore, this model represents a useful, easy way to correct data measured with GSS sensors to return "FRM-like" ozone concentrations by including two variables easily measurable (or available from existing monitoring networks).

Another model (model 3) was tested by adding NO_2 to model 2. Despite results indicating an improvement in the goodness of fit with respect to model 2 (adjusted r^2=0.91, RMSE=4.2 ppb and partial F-test p<0.001), model 3 also exhibited higher RMSE of the hourly- and weekly-averaged data (Figure 2) and a slight increase of the Theil-Sen slope (Figure S7). Model 3 results showed larger differences during weekends (particularly on Sundays) that are likely related to the decreased NO_2 concentrations with the lower road traffic emissions. It also confirms the work of Lin *et al*. (2015) showing that NO_2 had only a limited effect on the O_3 sensor.

3.6 Viability of proposed models

The main purpose in using low-cost monitors is to increase the spatial resolution of routine monitoring networks by deploying a larger number of sites. Since individual sensors may vary in performances, the 10 monitors were co-located under lab conditions and exposed to different O_3 concentrations (<5, 50, and 100 ppb). The inter-monitor Pearson correlation coefficients ranged from 0.97 to 0.99, indicating a very good agreement among the monitors. This result allows applying the proposed correction models to the multiple monitors deployed at locations other than where an FEM monitor is deployed.

A potential limitation of the study is the deployment of monitors over only 4 months in summer/early fall weather conditions. The ambient air temperatures recorded throughout the sampling campaign varied between 0°C and 38°C, allowing the extension of correction models over the sensor's operating temperature range (0°C to 40°C). The deployment of these GSS sensors outdoors under colder weather conditions would likely be possible when the monitors are placed in a warmed enclosure (e.g. using a small light bulb for heating). However, a careful evaluation of the validity of the proposed correction under such conditions must be performed.

4 Conclusions

The poor and sparse spatial coverage of current air quality monitoring networks in the U.S. limits the assessment of exposure to O_3 in epidemiological studies. Limited predictive accuracy may lead to potential exposure misclassification resulting in negatively biased health effect estimates. This study demonstrate that the Aeroqual Series 500 monitors coupled with O_3 semiconductor oxide sensors could provide improved spatial coverage in routine monitoring

networks. However, a careful calibration must be performed prior their deployment to check or adjust the original calibration. The monitor worked well over 4 months with no problems. The raw field data showed a ~12% overestimation bias with respect to data measured under clean laboratory air conditions. The application of correction models including the temperature in the monitor enclosure and ambient RH was able to produce "FRM-like" ozone concentrations explaining 90% of the variance.

ACKNOWLEDGEMENTS

This work was funded by the National Institute of Environmental Health Sciences (Grant #P30 ES001247). The authors gratefully acknowledge (i) the New York State Department of Environmental Conservation for providing the air quality data measured at the ROC site; (ii) NOAA National Climatic Data Center for providing weather data; and (iii) NOAA/OAR/ESRL PSD, Boulder, Colorado, USA for providing NCEP Reanalysis data from their Web site at https://www.esrl.noaa.gov/psd/.

REFERENCES

Aeroqual (2018). Aeroqual Series 200/300/500 User Guide. Aeroqual Ltd. Auckland, New Zealand. Available at: https://d2pwrbx99jwry6.cloudfront.net/wp-content/uploads/Environmental-factors-in-the-operation-of-GSS-ozone-sensors.pdf (last accessed January 2018)

Ainsworth, E.A., Yendrek, C.R., Sitch, S., Collins, W.J. and Emberson, L.D. (2012). The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annu. Rev. Plant Biol.*, 63: 637-661.

Aliwell, S.R., Halsall, J.F., Pratt, K.F.E., O'Sullivan, J., Jones, R.L., Cox, R.A., Utembe, S.R., Hansford, G.M. and Williams, D.E. (2001). Ozone sensors based on WO3: a model for sensor drift and a measurement correction method. *Meas. Sci. Technol.*, 12: 684-690.

Bart, M., Williams, D.E., Ainslie, B., McKendry, I., Salmond, J., Grange, S.K., Alavi-Shoshtari, M., Steyn, D. and Henshaw, G.S. (2014). High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia. *Environ. Sci. Technol.*, 48: 3970-3977.

Bell, M.L., Zanobetti, A. and Dominici, F. (2014). Who is more affected by ozone pollution? A systematic review and meta-analysis. *Am. J. Epidemiol.*, 180: 15-28.

Carslaw, D.C., Beevers, S.D., Ropkins, K. and Bell, M.C. (2006). Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. *Atmos. Environ.*, 40: 5424-5434.

Cooper, O.R., Parrish, D.D., Ziemke, J., Balashov, N.V., Cupeiro, M., Galbally, I.E., Gilge, S., Horowitz, L., Jensen, N.R., Lamarque, J.F. and Naik, V. (2014). Global distribution and

trends of tropospheric ozone: An observation-based review. *Elementa*, 2: 000029, doi: 10.12952/journal.elementa.000029

Emami, F., Masiol, M. and Hopke, P.K. (2018). Air pollution at Rochester, NY: Long-term trends and multivariate analysis of upwind SO2 source impacts. *Sci. Total Environ*., 612: 1506-1515.

Fowler, D., Pilegaard, K., Sutton, M.A., Ambus, P., Raivonen, M., Duyzer, J., Simpson, D., Fagerli, H., Fuzzi, S., Schjørring, J.K. and Granier, C. (2009). Atmospheric composition change: ecosystems–atmosphere interactions. *Atmos. Environ*., 43: 5193-5267.

Grange, S.K., Carslaw, D.C. and Lewis, A.C. (2016). Source apportionment advances with bivariate polar plots, correlation, and regression techniques. *Atmos. Environ*. 145: 128-134.

Jerrett, M., Burnett, R.T., Pope III, C.A., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E. and Thun, M. (2009). Long-term ozone exposure and mortality. *N. Engl. J. Med*., 360: 1085-1095.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L. and Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environ. Int*., 75: 199-205.

Lin, C., Gillespie, J., Schuder, M.D., Duberstein, W., Beverland, I.J. and Heal, M.R. (2015). Evaluation and calibration of Aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide. *Atmos. Environ*., 100: 111-116.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H. and Ek, M.B. (2006). North American regional reanalysis. *Bull. Am. Meteorol. Soc*., 87: 343-360.

Monks, P.S. (2005). Gas-phase radical chemistry in the troposphere. *Chem. Soc. Rev*., 34: 376-395.

Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E. and Stevenson, D.S. (2015). Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmos. Chem. Phys*., 15: 8889-8973.

NYSDEC (New York State Department of Environmental Conservation) (2017). 2017 Annual Monitoring Network Plan, New York State Ambient Air Monitoring Program. Bureau of Air Quality Surveillance, NYSDEC, Albany, NY, http://www.dec.ny.gov/docs/air_pdf/2017plan.pdf, Last access: 20 January 2018.

Palmer, P.I., Jacob, D.J., Fiore, A.M., Martin, R.V., Chance, K. and Kurosu, T.P. (2003). Mapping isoprene emissions over North America using formaldehyde column observations from space. *J. Geophys. Res.-Atmos*., 108: 4180, doi:10.1029/2002JD002153.

Seinfeld, J.H. and Pandis, S.N. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, Hoboken, NJ.

Sen, P.K. (1968). Estimates of regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc*., 63: 1379-1389.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J.F., Kuhn, U., Stefani, P. and Knorr, W. (2014). Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years. *Atmos. Chem. Phys*., 14: 9317-9341.

Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S., Shelow, D., Hindin, D.A., Kilaru, V.J., and Preuss, P.W. (2013). The changing paradigm of air pollution monitoring. *Environ. Sci. Technol*., 47: 11369–11377.

Stevenson, D.S., Young, P.J., Naik, V., Lamarque, J.F., Shindell, D.T., Voulgarakis, A., Skeie, R.B., Dalsoren, S.B., Myhre, G., Berntsen, T.K., and Folberth, G.A. (2013). Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmos. Chem. Phys*., 13: 3063-3085.

Theil, H. (1950). A rank invariant method of linear and polynomial regression analysis, i, ii, iii. *Proc. Konink. Nederl. Akad. Wetensch. A*., 53: 386-392, 521-525, 1397-1412.

Turner, M.C., Jerrett, M., Pope III, C.A., Krewski, D., Gapstur, S.M., Diver, W.R., Beckerman, B.S., Marshall, J.D., Su, J., Crouse, D.L. and Burnett, R.T. (2016). Long-term ozone exposure and mortality in a large prospective study. *Am. J. Respir. Crit. Care Med*., 193: 1134-1142.

Uria-Tellaetxe, I., and Carslaw D.C. (2014). Source identification using a conditional bivariate Probability function. *Environ. Mod. Softw*., 59: 1-9.

USEPA (United States Environmental Protection Agency) (2013). Transfer Standards For Calibration of Air Monitoring Analyzers for Ozone, Technical Assistance Document, Report No. EPA -454/B-13-004, Office of Air Quality Planning and Standards, Air Quality Assessment Division, https://www3.epa.gov/ttnamti1/files/ambient/qaqc/OzoneTransferStandardGuidance.pdf, Last access: 8 February, 2018.

USEPA (United States Environmental Protection Agency) (2017). List of Designated Reference and Equivalent Methods, Issue of June 16, 2017, USEPA National Exposure Research Laboratory, Triangle Park, NC. https://www3.epa.gov/ttn/amtic/criteria.html, Last access: 22 January 2018.

White, R.M., Paprotny, I., Doering, F., Cascio, W.E., Solomon, P.A. and Gundel, L.A. (2012). Sensors and 'apps' for community-based: Atmospheric monitoring. EM: *Air and Waste Management Association's Magazine for Environmental Managers*, (MAY), 36-40.

Williams, D.E., Henshaw, G.S., Wells, D.B., Ding, G., Wagner, J., Wright, B.E., Yung, Y.F. and Salmond, J.A. (2009). Development of low-cost ozone measurement instruments suitable for use in an air quality monitoring network. *Chem. New Zealand*, 73: 27–33

Williams, D.E., Henshaw, G.S., Bart, M., Laing, G., Wagner, J., Naisbitt, S. and Salmond, J.A. (2013). Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. *Meas. Sci. Technol*., 24: 065803.

TABLES

Table 1. Results of the regressions analyses to return "FEM-like" Aeroqual concentrations.

Selected model	Coefficients and estimators					
		Unit	Estimate	Std. error	t value	*p*
Calibration: Aeroqual=β_0+(β_1·FEM)						
	R^2	—	0.87	—	—	—
	β_0	ppb	8.64	0.211806	40.8	<0.001
	β_1	ppb	0.81	0.006041	134.8	<0.001
	RMSE	ppb	5.0	—	—	—
	MAE	ppb	3.9	—	—	—
Model 1: Aeroqual=β_0+(β_1·FEM)+(β_2·Encl.Temp.)+(β_3·RH)						
	R^2	—	0.89	—	—	—
	R_{adj}^2	—	0.89	—	—	—
	β_0	ppb	20.139	0.771	26.1	<0.001
	β_1	ppb	0.912	0.009	105.6	<0.001
	β_2	°C	-0.476	0.021	-22.4	<0.001
	β_3	%	-0.059	0.006	-9.4	<0.001
	RMSE	ppb	4.6	—	—	—
	MAE	ppb	3.5	—	—	—
Model 2: Aeroqual=β_0+(β_1·FEM)+(β_2·Encl.Temp.)+(β_3·Encl.Temp.^{2})+(β_4·RH)						
	R^2	—	0.90	—	—	—
	R_{adj}^2	—	0.90	—	—	—
	β_0	ppb	11.730	0.901	13.0	<0.001
	β_1	ppb	0.925	0.008	111.5	<0.001
	β_2	°C	0.536	0.066	8.2	<0.001
	β_3	°C	-0.025	0.002	-16.2	<0.001
	β_4	%	-0.079	0.006	-12.8	<0.001
	RMSE	ppb	4.4	—	—	—
	MAE	ppb	3.4	—	—	—
Model 3: Aeroqual=β_0+(β_1·FEM)+(β_2·Encl.Temp.)+(β_3·Encl.Temp.^{2})+(β_4·RH)+(β_5·NO_2)						
	R^2	—	0.91	—	—	—
	R_{adj}^2	—	0.91	—	—	—
	β_0	ppb	13.953	0.865	16.1	<0.001
	β_1	ppb	0.875	0.008	104.3	<0.001
	β_2	°C	0.516	0.062	8.3	<0.001
	β_3	°C	-0.023	0.001	-16.1	<0.001
	β_4	%	-0.061	0.006	-10.2	<0.001
	β_5	ppb	-0.415	0.024	-17.5	<0.001
	RMSE	ppb	4.2	—	—	—
	MAE	ppb	3.1	—	—	—

FIGURES

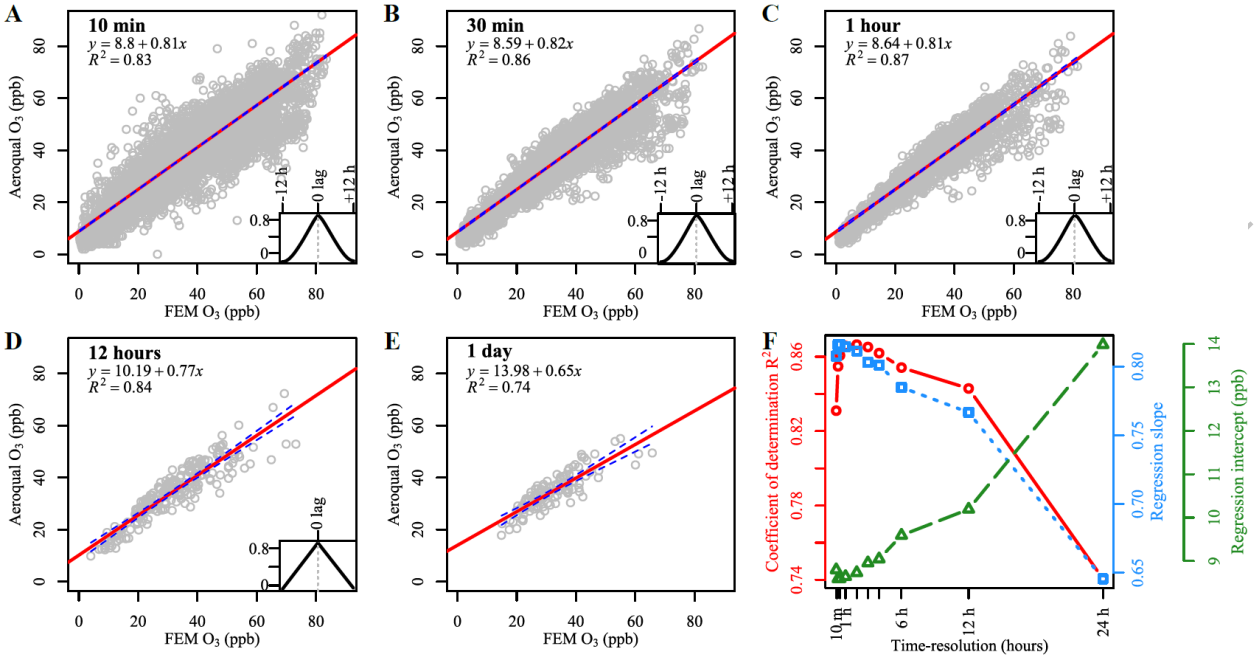

Figure 1. Linear regression of Aeroqual O_3 vs. FEM O_3 at varying time resolutions (a: 10 min; b:30 min; c: 1 h; d: 12 h; e: 24 h), and changes in the regression coefficients (f). The regression line is solid and red colored; the 95th confidence intervals are drawn in dashed blue. The regression plots (a-e) also report the cross-correlation curves (bottom-right).

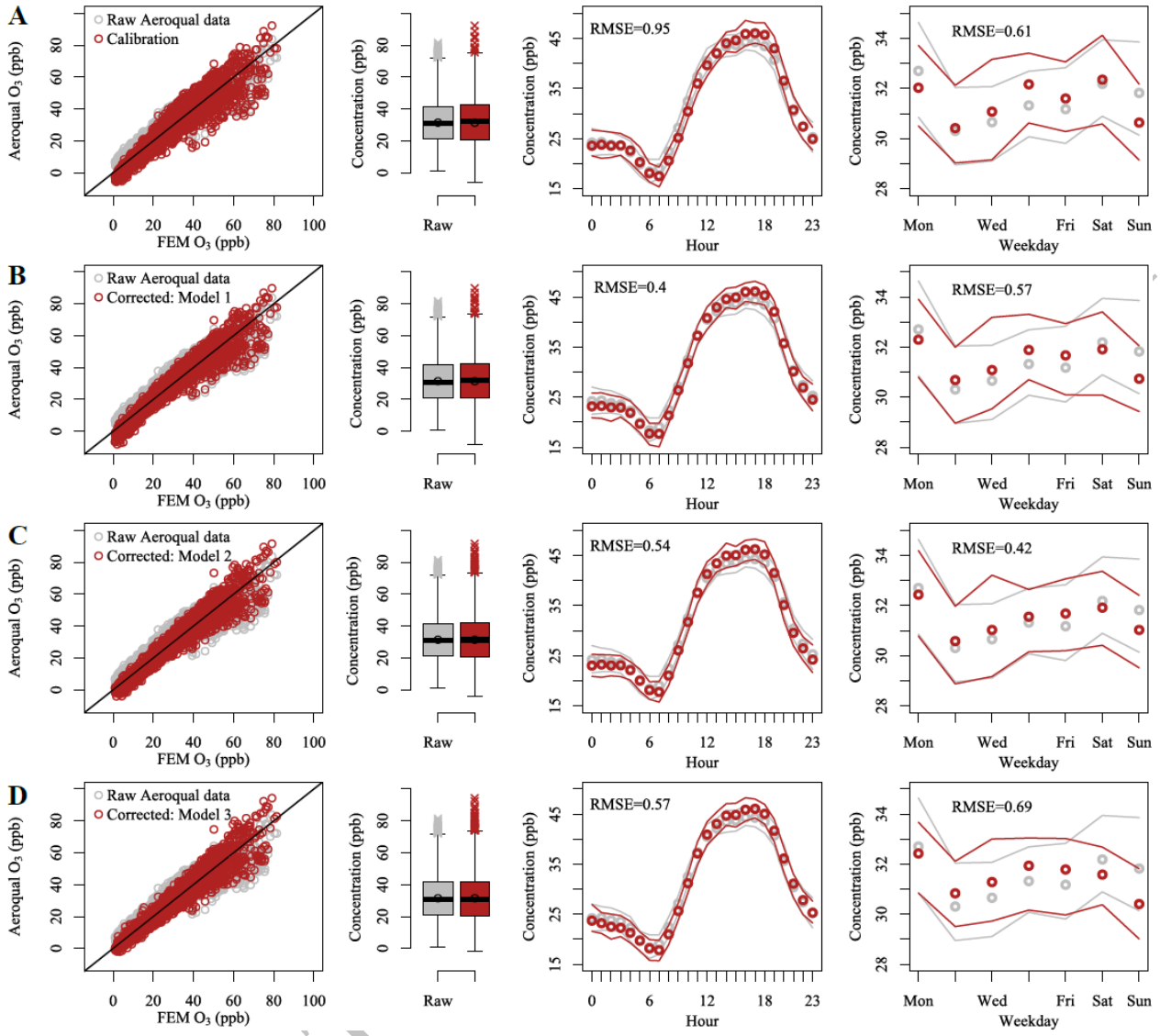

Figure 2. Results of the application of calibration and corrections. Each row of plots corresponds to a model, as in Table 1: calibration (A) and linear models (B= model 1, C= model 2, D= model 3). The scatterplots on the left show the fit of 1 h-averaged data with FEM (black lines represent the 1:1 relationship). The boxplots show the distribution of raw (grey) and corrected (red) data. Lines represent the medians, boxes are the 25-75th percentile ranges, whiskers show ±1.5*inter-quartile ranges, circles are the arithmetic means. The plots on the right show the diurnal and weekly patterns of raw (grey) and corrected (red) data. Circles represent the averages, lines draw the 95th confidence intervals of the averages computed by bootstrapping the data (n=2000). RMSE are also reported for comparing the goodness of the fits.