



Forecasting Surface O₃ in Texas Urban Areas Using Random Forest and Generalized Additive Models

Rick Pernak*, Matthew Alvarado, Chantelle Lonsdale, Marikate Mountain, Jennifer Hegarty, Thomas Nehr Korn

Atmospheric and Environmental Research, Lexington, MA 02421, USA

ABSTRACT

We developed and evaluated three types of statistical forecasting models (quantitative, probabilistic, and classification) for predicting the maximum daily 8-hour average concentration of ozone based on meteorological and ozone monitoring data for six Texas urban areas from 2009 to 2015. The quantitative and probabilistic forecasting models were generalized additive models (GAMs), whereas the classification forecast used the random forest machine learning method. We found that for the quantitative forecasting models, five of the eight predictors (the day of week, day of the year, water vapor density, wind speed, and previous day's ozone measurement) were significant at the $\alpha = 0.001$ level for all urban areas, whereas the other three varied in significance according to the location. The quantitative forecasting for the 2016 ozone season agreed well with the associated measurements (R^2 of ~0.70), but it tended to under-predict the ozone level for the days with the highest concentrations. By contrast, the probabilistic forecasting models showed little accuracy in determining the probability of concentrations exceeding policy-relevant thresholds during this season. The success rate for the random forest classification models typically exceeded 75% and would likely increase if the training data sets contained more extreme events.

Keywords: Ozone MDA8; Ozone prediction; Generalized additive models; Random forest.

INTRODUCTION

Ozone (O₃) at the surface can have adverse effects on public health. Lippmann (1989) summarized many studies that suggest lungs in people age quicker, lung capacity diminishes, and air flow resistance increases with sustained exposure to ozone. Devlin *et al.* (1997) examined lung inflammation and changes in lung function and found that these afflictions are likely due to elevated ozone and that those with pre-existing respiratory conditions are more prone to experience them. Studies performed on data from the United States, Canada, and Europe have linked air pollution to chronic obstructive pulmonary disease (COPD), increased hospitalization for respiratory illnesses (e.g., asthma), and lower forced expiratory volume (FEV). Anderson *et al.* (1997) confirmed an association with ozone and other pollutants with COPD in six European cities with different climates. Sixteen Canadian cities are the subject of Burnett *et al.* (1997), and they found that, in a population of over 12 million people over more than ten years, even low levels of surface

O₃ can lead to an increase in hospitalizations due to respiratory diseases. FEV was found by Brown *et al.* (2008) to decrease a statistically significant amount among young healthy adults after long exposures to ozone in the 60–80 ppb range. Finally, Bell *et al.* (2004) studied data that spanned over a decade for 95 US cities (and 40% of the population) and show that mortality rates increase a small but statistically significant amount when O₃ levels are raised by just 10 ppb. The EPA provides many more references on the effects of ozone on public health on their website (<https://www.epa.gov/ozone-pollution-and-your-patients-health/references-ozone-and-your-patients-health>) and in the Integrated Science Assessment (2013).

Given the harmful side effects of ozone inhalation, it is imperative that people know when to expect poor air quality days due to elevated ozone and what to do in such cases. State and local agencies across the country in conjunction with the EPA and AirNow recognize “Action Days” (<https://www.airnow.gov/index.cfm?action=airnow.actiondays>) when ozone and particulate matter concentrations are expected to be high. AirNow provides suggestions for residents of these areas to reduce the amount of pollutants in the air on these Action Days (<https://www.airnow.gov/index.cfm?action=resources.whatyoucando>), and the EPA provides services like the Air Quality Alert Program so that individuals who want poor air quality alerts can be emailed or notified on their

* Corresponding author.

Tel.: 781-761-2317; Fax: 781-761-2299
E-mail address: rpernak@aer.com

phone regarding a poor air quality day (<https://www3.epa.gov/region1/airquality/smogalrt.html>). State and local governments can also issue alerts via radio and television outlets so that individuals can decide to stay in their homes and reduce exposure to pollution. These kinds of precautions require accurate forecasting of air quality, so it is advantageous to have an ability to predict tropospheric O₃ concentrations so that sufficient precautions can be implemented. O₃ production is dependent on meteorological and photochemical conditions, so knowledge of both is necessary for accurate predictions.

Numerical models exist for this purpose, but statistical modeling techniques have been attempted to forecast surface O₃ as well. For example, Thompson *et al.* (2001) reviewed statistical methods for modeling the dependence of surface level O₃ on meteorology and emissions to obtain air quality forecasts, estimate O₃ time trends, and increase understanding of the underlying photochemistry that is responsible for the generation of ozone. These statistical methods were classified as regression (linear, tree-based, and non-linear), extreme value, and spatio-temporal modeling, none of which were found to be more appropriate than the others. Thompson *et al.* (2001) suggest that the best type of model to use is dependent on the region being considered and the machine learning technique employed.

Most of the literature focuses on forecasting modeling application to urban areas, with perhaps a similar application to rural areas as a baseline. For example, Feister and Balzer (1991) developed non-linear forecasting models for five (mostly non-urban and non-industrial) stations in Germany using over 300 predictors and found that recent O₃ surface levels and solar irradiance were the most significant but explained no more than 46% of the variance. Camalier *et al.* (2007) determined the effect of meteorology on O₃ in 39 urban areas in the eastern United States using a generalized linear model and found that ozone concentration increases with the temperature and decreases with increasing relative humidity. They also show that the temperature dependence is more pronounced in the cities in their study at higher latitudes, while humidity is the more significant predictor at lower latitudes. Kgabi and Sehloho (2012) investigated the effect of both emissions and meteorology by monitoring surface O₃ in a rural area (Botsalano Game Reserve near the Botswana-South Africa border) and an industrial area (Marikana, South Africa) in Africa and measuring the associated temperature, relative humidity, wind speed, and wind direction. They observed consistently higher ozone concentrations at the industrial location, a negative correlation between O₃ and humidity, and a positive correlation between temperature and O₃. More recently, statistical models have been used to identify unusual conditions, such as wildfires, that may have contributed to ozone formation. For example, Gong *et al.* (2017) developed a Generalized Additive Model (GAM) approach to estimate O₃ in several cities and then showed that days with measurable fire smoke impact tended to have positive residuals, suggesting the fires were contributing to O₃ formation on these days.

In addition to the GAM approach, other statistical forecasts for surface O₃ have used artificial neural networks, or ANNs (Abdul-Wahab and Al-Alawi, 2002; Luna *et al.*, 2014), and

random forests. ANNs are non-linear in nature, and the relationships that are discovered are deduced with limited prior knowledge of a given process. They are thus helpful in forecasting ozone for particular locations and provide the ability to model the change in O₃ concentrations due to meteorology and photochemical processes together (by including concentrations of pollutants, specifically primary pollutants like NO_x and NMHC, as predictors in the model). Abdul-Wahab and Al-Alawi (2002) implemented this procedure for Kuwait, while Luna *et al.* (2014) applied it to Rio de Janeiro in Brazil. Both studies produced ozone predictions that yielded > 90% correlation with measurements in training and testing, thus providing models that account for nearly all of the variation in the observations. Abdul-Wahab and Al-Alawi (2002) found that meteorology explains up to 41% of the variance in their models, while variability in the emissions accounted for the rest of the variability in O₃ concentrations. Luna *et al.* (2014) included CO, NO, moisture content, and NO_x as inputs into their ANN model and showed that the statistical model produced forecasts that are consistent with our knowledge of the chemical processes of ozone production.

Random forests are classifying algorithms that utilize a large number of independent, identically distributed (i.i.d.) decision trees with randomly selected predictors and average the results to reduce noise (Breiman, 2001; Hastie *et al.*, 2017). Yu *et al.* (2016) apply this technique to categorize air quality conditions in Shenyang, China, using the Chinese AQI and found that their random forest algorithm (RAQ) was more precise and accurate than other machine learning classification methods (Naïve Bayes, Logistic, etc.), though their application was not necessarily a forecast of ozone levels since the Chinese AQI is a maximum value for a number of pollutants (SO₂, NO₂, CO, PM_{2.5}, PM₁₀, and O₃).

In this work, we developed statistical forecasting models of surface O₃ for six urban areas in Texas (Table 1). We focused on forecasting ozone during eastern Texas's "ozone season" of May through October. We used generalized additive models—GAMs (Wood, 2006)—to describe the potentially non-linear relationship between the measured maximum daily 8-hour average ozone concentrations ($O_{3,MDA8}$), selected meteorological variables from weather forecasts, and measurements of the previous day's surface O₃ levels based on data for 2009 to 2015. We developed two GAM-based models, a quantitative forecast that predicts the numerical value of $O_{3,MDA8}$ and a probabilistic model that predicts the probability that $O_{3,MDA8}$ will exceed a given threshold. We then used the random forest method to develop a classification forecast of the O₃ air quality index (AQI)—to our knowledge this is the first time this method has been used to forecast surface O₃ concentrations directly. Here we describe our model development and the data used to train and evaluate the models. We then assess the performance of the models in forecasting $O_{3,MDA8}$ during the 2016 O₃ season in the urban areas of interest.

METHODS

Training and Evaluation Data

To develop the quantitative and probabilistic forecasting

models, we fitted GAMs to calculate the predicted $O_{3,MDA8}$ for day D_i in each urban area listed in Table 1 using ozone season data from 2009 to 2015, with the day D_{i-1} $O_{3,MDA8}$ calculated from the Texas Commission on Environmental Quality (TCEQ) monitoring data as a predictor. The other predictors used in these models are shown in Table 2. $O_{3,MDA8}$ values from the Texas Air Monitoring Information System (TAMIS) for D_{i-1} were the only measurements that were used in the forecasting model (Available at <https://doi.org/10.5281/zenodo.2032330>). We also used seven forecasted (D_i) meteorological parameters from the Model Output Statistics (MOS) post-processing of the National Center for Environmental Prediction (NCEP) 12-km North American Model (NAM-12) forecasts (Available at <https://doi.org/10.5281/zenodo.1979000>). We discuss the selected predictors for each data set further in this section.

Texas Commission on Environmental Quality Monitoring Data

TCEQ provided air quality data from the air quality monitoring network operated by TCEQ, its grantees, or local agencies whose data are stored in the TAMIS in and near the urban areas listed in Table 1. Historical data for the time period spanning 2009–2015 were provided by TCEQ and used in the training of the forecasting models. We used these data to calculate $O_{3,MDA8}$ for a given day over all sites around an urban area with acceptable data. For the 2016 ozone season evaluation period, the operational forecasting models gathered these $O_{3,MDA8}$ data in near-real time (NRT) from the TCEQ website (https://www.tceq.texas.gov/cgi-bin/compliance/monops/daily_average.pl), which provided raw hourly measurements of O_3 from all of the monitoring stations in its network. Note that while the training data were subjected to a quality control process by TCEQ, this was not true of the NRT data used in the operational forecasts. These processing differences had a small effect on our forecasts. For example,

the mean and standard deviation for the NRT Austin $O_{3,MDA8}$ data were 43.17 and 9.68 ppbv, respectively, while the moments of the corresponding quality-controlled distribution were 44.46 and 9.95 ppbv.

North American Mesoscale Numerical Forecasts

NCEP provides daily output from the 12-km (NAM-12) weather forecasting model. For a given model runtime, there are a number of forecast hours and grid types, but the product most relevant for this application is the one produced for the continental United States on a 12-km Lambert-conic conformal grid. Every 6 hours, model output is written to a single file and archived in the NOAA Operational Model Archive and Distribution System (NOMADS; Rutledge *et al.*, 2006), which is available to the public (<http://nomads.ncep.noaa.gov/pub/data/nccf/com/nam/prod>). Historical data (data set ds609.0) and forecasts (ds335.0) are archived in the Research Data Archive (RDA; NCEP/NWS/NOAA/ECMWF/UCAR, 2003) at the University Cooperation for Atmospheric Research (UCAR) (<https://rda.ucar.edu/>). Both the NRT and historical data were processed in the same way: Grid points closest to the coordinates of the areas of interest (Table 1) were determined; then the temperature, relative humidity, and geopotential height fields were extracted for the 500, 700, 850, and 925 mbar pressure levels. As mentioned, there are four forecast times (00:00, 06:00, 12:00, and 18:00 UTC), but only the 12Z forecast was used in this work, as this was the time found to give the best $O_{3,MDA8}$ fit for cities in the eastern U.S. by Camalier *et al.* (2007). This forecast time is at 07:00 CDT for our urban areas during the ozone season.

Model Output Statistics Forecasts

Model output statistics (MOS) is a technique used by the National Weather Service (NWS) to objectively interpret numerical model output (Carter *et al.*, 1989) and produce site-specific guidance (Glahn *et al.*, 2008). Archived and

Table 1. TCEQ and MOS sites corresponding to the six urban areas investigated in this report.

TCEQ Area Abbreviation	MOS Site Abbreviation	MOS Site Location (latitude, longitude)	Associated Cities
ARR	KAUS	30.321, -90.760	Austin/Round Rock
BPA	KBPT	29.951, -94.021	Beaumont/Port Arthur
DFW	KDFW	32.898, -97.019	Dallas/Fort Worth
HGB	KIAH	29.980, -95.360	Houston/Galveston/Brazoria
SA	KSAT	29.544, -98.484	San Antonio
TLM	KGGG	32.385, -94.712	Tyler/Longview/Marshall

Table 2. Predictors used in the GAM-based forecasts (quantitative and probabilistic) and random forest classification forecast.

Predictor	Units
Day of week (D_w)	None
Day of year (D_i)	None
MOS D_i Forecast Afternoon (21:00 UTC) Temperature (T_{2100})	°C
MOS D_i Forecast Diurnal Temperature Difference (ΔT)	°C
MOS D_i Forecast Daily Average water vapor density (ρ_{wv})	g m^{-3}
MOS D_i Forecast Daily Average Wind Speed (W_S)	m s^{-1}
MOS D_i Forecast Daily Average Wind Direction (W_D)	degrees clockwise from north
TAMIS O_3 MDA8 for D_{i-1} ($O_{3,MDA8}$)	ppbv

NRT-forecasted parameters that were used in this study include temperature, dew point, wind speed, and wind direction. The forecasts are provided every 3 hours for the subsequent ~2.5 days. All are calculated for the same day (D_i) relative to the model runtime (which is at midnight). MOS forecasts are archived by the Iowa Environmental Mesonet (IEM) at Iowa State University (<https://mesonet.agron.iastate.edu/mos/fe.phtml>)—the archived NAM-12 forecasts were used in the model training for this study. NRT NAM-12 forecasts are provided by NWS (<http://www.nws.noaa.gov/mdl/synop/products/bullform.met.php#Texas>) and were used for forecasting. Queries were performed for the MOS sites listed in Table 1.

National Air Quality Forecast System

The National Air Quality Forecast System (NAQFC; Pan et al., 2014) is a physical modeling system provided by NOAA that couples the Weather Research and Forecasting Non-hydrostatic Mesoscale Model (WRF-NMM) (Janjic, 2003) and the Community Multiscale Air Quality (CMAQ) (Byun and Schere, 2006) regional chemical transport model. CMAQ is a multi-scale 3D Eulerian chemical transport model that is used to model air quality (e.g., tropospheric O_3 and $PM_{2.5}$) on urban to regional scales. During this study, the CB05 chemical mechanism (Yarwood et al., 2005) and AERO5 aerosol module (Carlton et al., 2010) are used in CMAQ v4.7.1 within the NAQFC system. The lateral boundary conditions used in the simulation are monthly averaged profiles extracted from GEOS-Chem (Bey et al., 2001) simulation results. The current operational model has a horizontal resolution of 12 km. The O_3 forecasts are produced twice daily from the 06:00 and 12:00 UTC cycles. Forecast time projections extend from a minimum of 30 hours (06:00 run) to a maximum of 48 hours (12:00 run). Note that NOAA provided their forecast output for a subset of our testing period, 19 May 2016 through 22 September 2016.

Forecasting Models

Quantitative and Probabilistic Models

In these two procedures, we used the “mgcv” GAM modeling library (version 1.8–23) in R (version 3.2.2) discussed in Wood (2006) to fit the $O_{3,MDA8}$ value for day D_i using a number of meteorological and air quality parameters as predictors. The general GAM formulation used in our forecasting models can be written as follows:

$$g(\mu_i) = \beta_0 + f_1(x_{i,1}) + f_2(x_{i,2}) + \dots + f_n(x_{i,n}) + f_p(D_i) + D_W \quad (1)$$

where $g(\mu_i)$ is the expected value of $O_{3,MDA8}$ for day i in the quantitative models and a Boolean value indicates whether the $O_{3,MDA8}$ exceeded a given threshold in the probabilistic models. $g(\mu_i)$ is the link function: A log-link is used for the quantitative forecasting models, and a logistic link is used for the probabilistic forecasting models. The $j = 1, \dots, n$ meteorological and air quality parameters are denoted with $x_{i,j}$, with the corresponding $f_j(x_{i,j})$ being an initially unknown smooth function of $x_{i,j}$ made from a cubic-spline basis set. Following Camalier et al. (2007), two other predictors are also included: a smooth function $f_p(D_i)$ of the Julian day of

the year (D_i) and a factor for the day of the week D_W . As we are only fitting O_3 data during the O_3 season (May–October), $f_p(D_i)$ is built with a non-periodic cubic spline basis.

Random Forest O_3 Classification Forecasts

We also used the random forest approach to develop a third forecasting model. Bosch et al. (2007) and Hastie et al. (2017) provide detailed documentation on how decision trees and their associated random forests operate (Also see slides 25–36 in <http://www.robots.ox.ac.uk/~az/lectures/ml/lect4.pdf>), which we will only summarize. Given a set of data and parameters that will be used in event classification, a set of true–false tests are performed (tree nodes), and the end result is a number of end nodes (leaves, where the tests stop) that yield a percentage of each class that each leaf contains. The set of parameters that defines a class is the one that yields the highest probability of a given class. Extending this methodology to a number of trees—each of which contains a random subset of the training data and performs node tests on a random subset of the given predictors—is what is known as a random forest, and it reduces the risk of over-fitting. For the forest, the probability of a given class given a set of parameters is averaged over the ensemble of trees, and class designation is again defined by a maximum probability (this time, over the ensemble average). For random forest implementation in this study, we used the same parameters that were selected for the quantitative and probabilistic forecasting models (Table 2) and the color-coded classification scheme for O_3 air quality indices (AQIs) in Table 3.

Predictor Selection

We performed several sensitivity studies that assessed model performance for different combinations of predictors with other data sources. The sensitivity studies included:

- a base model with the parameters listed in Table 2;
- replacing the $O_{3,MDA8}$ derived from TCEQ measurements with the EPA AirNow AQI;
- utilizing a stability parameter— ΔT_{NCEP} —defined as the difference between the NCEP-predicted temperature at 925 mbar, T_{925} , and the temperature at 850 mbar, T_{850} ;
- using relative humidity or dew point temperature as a proxy for water vapor density;
- replacing MOS and NCEP forecasts with TCEQ current conditions;
- using HYSPLIT 24-hour back trajectory distances and bearings in addition to the TCEQ, MOS, and NCEP parameters; and

Table 3. $O_{3,MDA8}$ classes used in the classification model. The ranges are defined by the EPA and summarized by the TCEQ at https://www.tceq.texas.gov/cgi-bin/compliance/monops/ozone_summary.pl#interpret.

Color Class	$O_{3,MDA8}$ Range (ppb)
Green	$O_{3,MDA8} < 55$
Yellow	$55 \leq O_{3,MDA8} < 71$
Orange	$71 \leq O_{3,MDA8} < 86$
Red	$86 \leq O_{3,MDA8} < 105$
Maroon	$105 < O_{3,MDA8}$

- using HYSPLIT-deduced synoptic types instead of the bearings and distances.

Ultimately, each sensitivity study produced statistics—best-fit line slope and intercept, Pearson r correlation coefficient, and mean and standard deviation of residuals—that were generally worse than the model described in Table 2 when validated with data from the evaluation period.

ΔT_{NCEP} , the difference in NCEP-forecasted temperature between 925 and 850 mbar, was only significant to the 5% level for Austin and San Antonio; thus, with this predictor, over-fitting of the data was a potential danger. In addition, by omitting this parameter, NCEP forecast gathering was no longer necessary. An added advantage with this improvement was that the forecaster became much more efficient—NCEP processing took approximately 40 seconds per city, so excluding this data set saved 4 minutes of runtime per forecast.

The moisture-related MOS-forecasted parameters H_R (relative humidity) and T_{DP} (dew point temperature) were attempted before water vapor density because of their usage in Camalier *et al.* (2007), and they were usually significant to at least the 1% level. The significance level was dependent on the urban site, particularly when T_{DP} and H_R were used together. Fit separately, T_{DP} is a significant predictor at the 0.1% level for all sites except DFW, while H_R is significant to the highest level (0.1%) for all locations. We also investigated how an absolute measure of moisture could affect the predictions by transforming these variables into water vapor density (ρ_{WV}). Like H_R , ρ_{WV} is significant to the highest level (0.1%) for all locations. All of these significances are summarized in Supplementary Table 1. Water vapor density better reflects the impact water vapor can have on the chemical production of ozone than relative humidity or dew point temperature (as chemical reactions depend on concentrations in units of molecules cm^{-3}), though, and consequently is the preferred moisture predictor. H_R and T_{DP} were thus removed from our model and replaced with an average water vapor density that is a function of temperature (in $^{\circ}\text{C}$) and relative humidity. ρ_{WV} is computed every hour between 06:00 UTC and 21 hours later, and then an average ρ_{WV} is found.

RESULTS AND DISCUSSION

To assess the performance of the forecasting models, we compared the observed and predicted $O_{3,MDA8}$ for the 2009–2015 training data set and for the 2016 ozone season. Assessing accuracy of the models in the training period is important when, for example, determining whether models should be applied in an operational environment to compute NRT predictions of O_3 . If the residuals do not follow a normal distribution, the fitting needs to be refined. To quantify model accuracy completely, though, one must calculate the O_3 level predictions and then compare them with associated measurements.

Evaluating the Statistical Fit

Quantitative Forecasts

Table 4 presents the R^2 correlation coefficient for the quantitative forecasting models for each urban area. The models typically can explain over 70% of the variance ($R^2 >$

Table 4. R^2 and standard deviation (σ) of the residuals for the quantitative model training.

Urban Area	R^2	σ of Residuals (ppbv)
ARR	0.755	6.61
BPA	0.726	8.36
DFW	0.716	8.57
HGB	0.734	9.77
SA	0.734	7.31
TLM	0.698	7.72

0.7), with the best fits found for Austin/Round Rock and the worst for Tyler/Longview/Marshall (the only model that does not explain at least 70% of the variance). Examination of the quantitative forecasting model residuals shows that the fits to the training data set are generally good, with a normal distribution of residuals, little trend in the residuals with the magnitude of the predicted or fitted values, and a linear relationship between the measured and fitted $O_{3,MDA8}$. An example is provided in Fig. 1 for the Austin/Round Rock quantitative forecasting model, which plots the quantile-quantile (Q-Q) comparison, residuals as a function of the log of predictor value, residual distribution, and modeled $O_{3,MDA8}$ as a function of measured $O_{3,MDA8}$. Additionally, Fig. 2 exhibits the functional dependence of the $O_{3,MDA8}$ predictions with each predictor. As expected, the forecasted $O_{3,MDA8}$ increases as the previous day's $O_{3,MDA8}$ increases and as the forecasted afternoon temperature increases, but decreases as water vapor density and wind speed increase.

Supplementary Table S1 shows the estimated significance of each predictor in the quantitative forecasting models. We see that five of the eight predictors are significant for all urban areas to the $\alpha = 0.001$ level. The remaining predictors are forecasted afternoon temperature, forecasted daily temperature difference, and average forecasted wind direction. The significance of each depends on the city, but even in most of these cases, the predictors are significant to at least the $\alpha = 0.05$ (afternoon temperature) or $\alpha = 0.10$ (diurnal temperature difference and wind direction) level. Wind direction at TLM is the one exception. p -values for separate runs with T_{DP} , H_R , T_{DP} and H_R together, ΔT_{NCEP} , and ρ_{WV} demonstrate that no significance is lost when omitting ΔT_{NCEP} and T_{DP} and that H_R and ρ_{WV} (both of which are derived from the MOS T_{DP} and surface temperatures) are equally significant (i.e., to the $\alpha = 0.001$ level) in predicting ozone levels. In selecting a moisture predictor, we decided to use ρ_{WV} , which yielded a substantially higher correlation coefficient for Austin, Beaumont/Port Arthur, and Houston.

Probabilistic Forecasts

We trained probabilistic forecasting models that determine the odds of predicted $O_{3,MDA8}$ being greater than or equal to four thresholds—55 ppb, 71 ppb, 86 ppb, and 105 ppb—given the MOS forecasts and the previous day's $O_{3,MDA8}$. However, most urban areas did not have sufficient historical data to fit GAMs for the upper two thresholds, and none had enough data to fit a 105-ppb model. Thus, most urban areas only have 55-ppb and 71-ppb probabilistic forecasting models, while DFW and HGB also have an 86-ppb model.

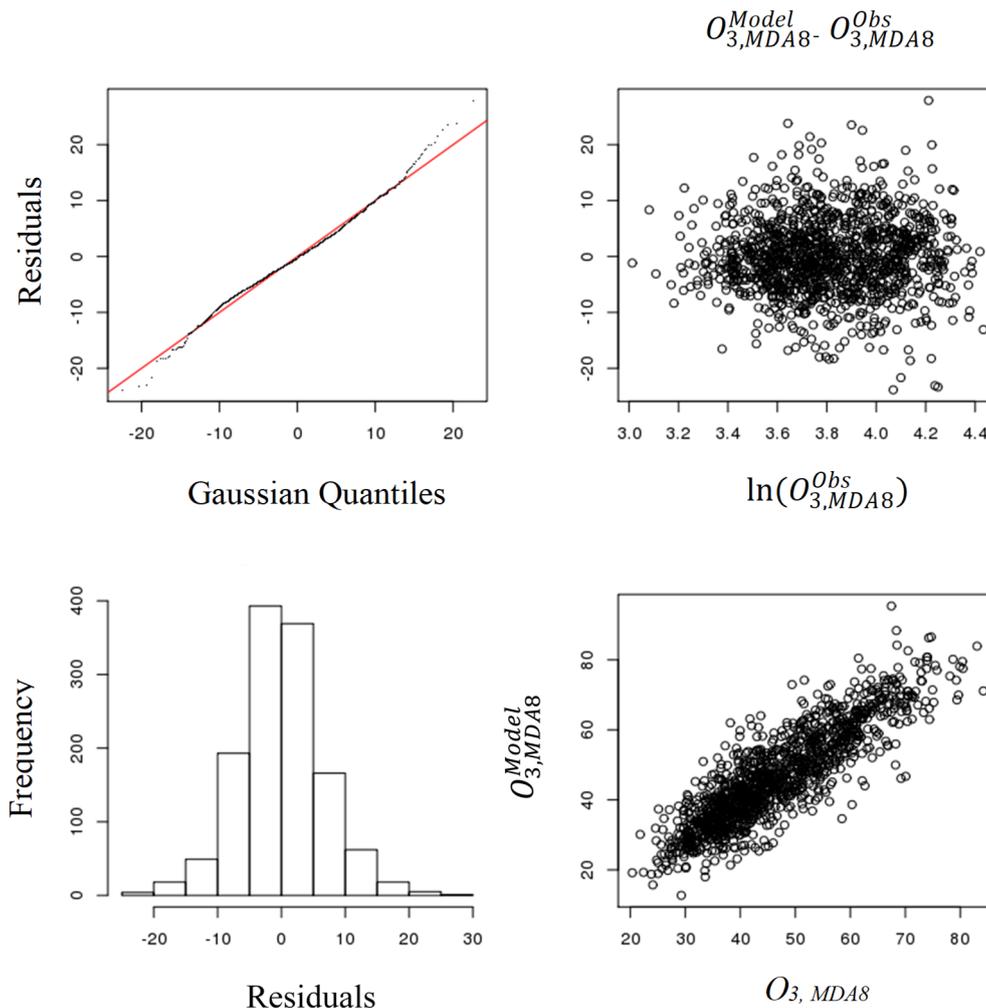


Fig. 1. Training period residual analysis for the Austin/Round Rock quantitative model. The top-left panel is a Q-Q plot which implies a normal distribution if the data points exhibit a $y = x$ relationship (represented by the red line). The top-right panel plots the data-model residuals as a function of the log of the measurements. A residual distribution is presented in the bottom-left panel. Finally, model predictions as a function of their associated measurements are plotted in the bottom-right panel. For ARR, the plots show that the quantitative model produces residuals that are normally distributed.

Supplementary Table S2 shows the R^2 and percentage variance explained for the probabilistic forecasting models. In general, the explained variance was relatively low for the probabilistic models, with values between 45% and 60%. Most, but not all, predictors were significant to at least the $\alpha = 0.10$ level, as shown in Supplementary Table S3. There is much more variance in predictor significance with the probabilistic models—the p -value changes with not only the urban area but the threshold level—compared to the quantitative models, whose predictors were typically significant to the $\alpha = 0.001$ level.

Examining the model fits shows some interesting differences between the urban areas. For example, the left panel of Fig. 3 shows results for HGB for a threshold of 71 ppb. Only the dependence of $O_{3,MDA8}$ on T_{2100} , and W_S are shown, with all other variables held constant at their mean values. These $O_{3,MDA8}$ forecasts are primarily a function of forecasted afternoon temperature (T_{2100}), showing a strong increase in the probabilities between 23°C and 35°C. However, the results

for SA in the right panel of Fig. 3 show that in this area, $O_{3,MDA8}$ and forecasted W_S are more important in predicting when the $O_{3,MDA8}$ will exceed 71 ppb.

O₃ Classification Forecasts

Accuracy in classification models can be computed by utilizing the associated confusion matrix, which is a table of what the forecaster predicted compared to what was observed (Kuhn, 2008). The diagonal elements of the matrix are successful classifications, and accuracy is given by the ratio of the number of successes to the total number of events. The sample size for the training period is 1280, and the success rate for each site during the training set is given in Table 5. Success rates for our classification models range between 65% and 85%, depending on the urban area of interest.

Forecasting Performance

Quantitative Forecasts

The quantitative models were used to forecast the entire

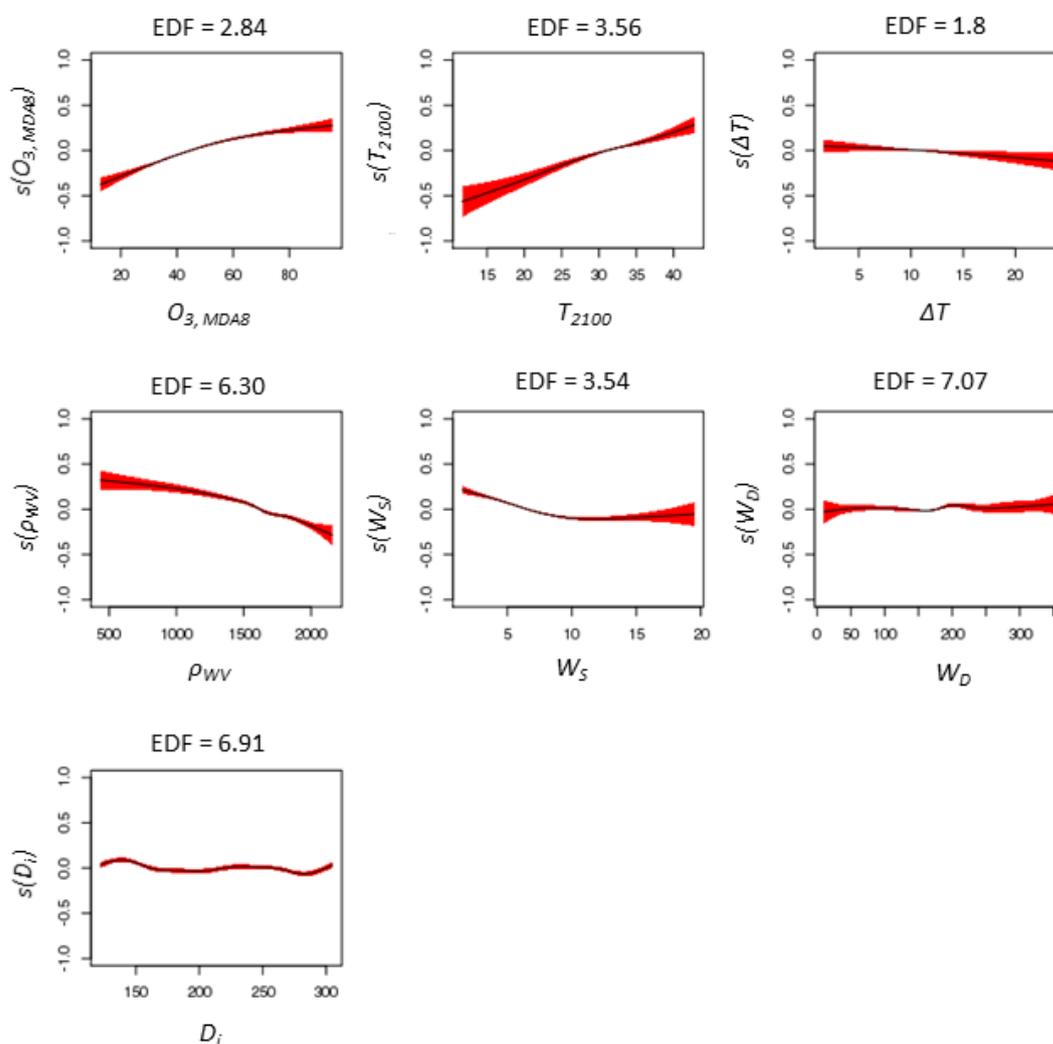


Fig. 2. Effects (i.e., weights) of each predictor as a function of predictor value, as represented by smoothed curves determined in the GAM fitting. Effective degrees of freedom (EDF) for each of the smooth functions is given in the title of each panel. 95% confidence levels are displayed as the red regions around the black curves. As an example, one interpretation of the T_{2100} plot is that its effect on the GAM response (predicted O_3) is estimated by a smooth curve with 3.56 degrees of freedom (Wood, 2006).

2016 evaluation period and then validated with the observations from TCEQ monitoring stations. Statistics to specify the forecast-measurement agreement were then calculated—these included the coefficients of the linear ordinary least-squares (OLS) fit (slope, a , and intercept, b), Pearson linear correlation coefficient (r), and the moments of the forecast-data residuals (mean bias, μ , and standard deviation, σ). These statistics are provided in Table 6. While not a one-to-one correspondence with the observations, the models do exhibit some predictive power, given that the correlation coefficients for all cities are on the order of 0.7 or higher.

We emphasize the slopes of the fits, all of which are less than unity, suggesting that the higher $O_{3,MDA8}$ measurements are underestimated and the lower measurements are overestimated in the models. This behavior is an artifact of the regression model—expectation values will be biased toward the mean because the mean minimizes the loss error

in the fit to the training data. However, there are a few reasons this kind of disagreement might be expected. As we alluded to in the TCEQ monitoring data training set section, NRT and training data were not subject to identical quality control protocol, and the training data are provided as floating point numbers, while the NRT data are given in integers. While these two quality control procedures will at most only account for an error of a couple ppb at a given data point, they can contribute to some of the noise that is exhibited in the validation statistics (e.g., correlation coefficient $r < 1$).

Another possible and probably more significant factor is a change in emissions, which is not accounted for in our models. With our training set, we assume that year-to-year changes in emissions and other variables are small enough throughout the training period such that the year was not used as a predictor. Furthermore, we assume that the emissions during the 2016 testing period resemble those from the

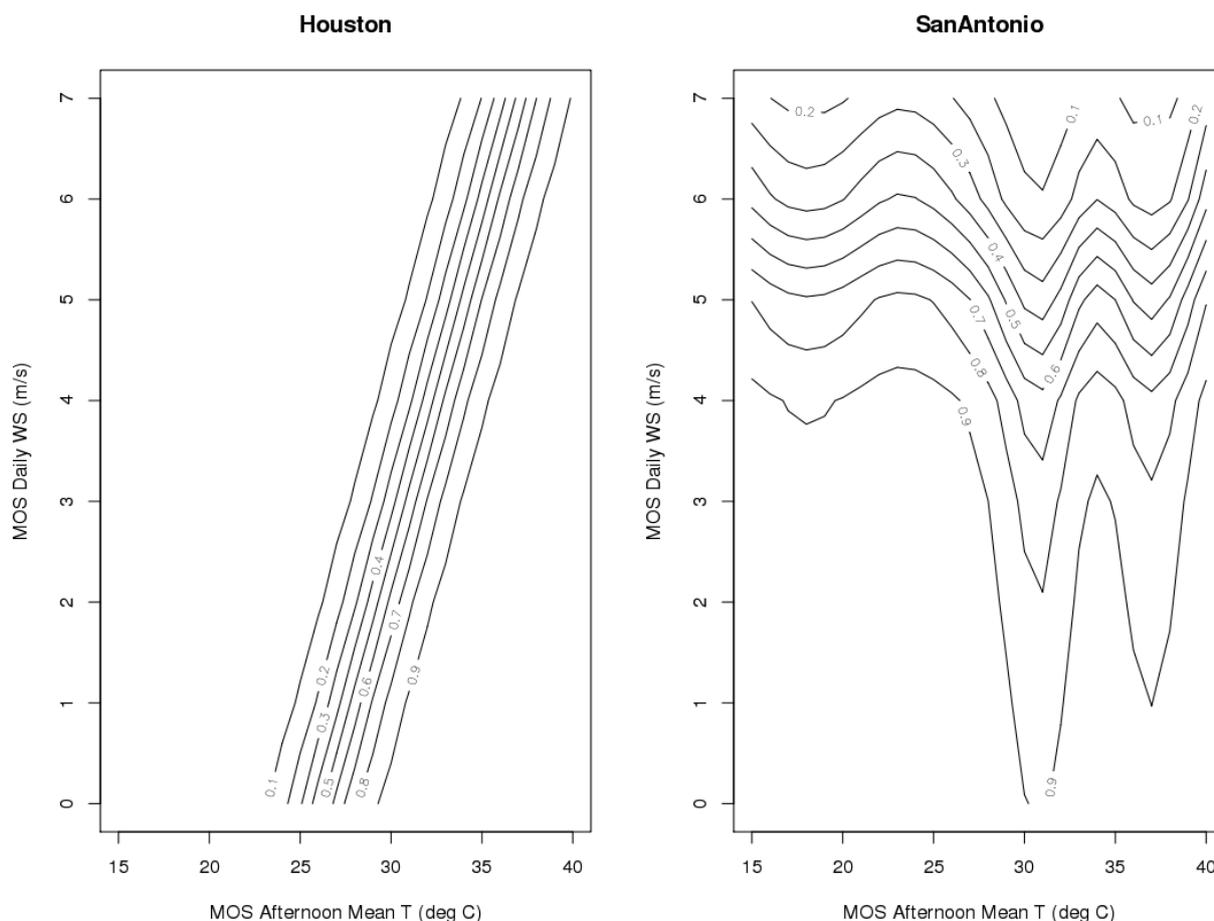


Fig. 3. Plot of the probabilities of forecasted $O_{3,MDA}$ being ≥ 71 ppb in Houston-Galveston-Brazoria (HGB) and San Antonio (SA) when the $O_{3,MDA}$ from the previous day is 70 ppb. The x -axis is the MOS NAM-12 forecast of afternoon mean temperature for a given day while the y -axis is the MOS NAM-12 forecast of daily averaged wind speed for a given day. All other model predictors are held constant at their mean values.

Table 5. Success rate for the classification models executed on the training data set.

City	Success Rate
Austin	84.4%
Beaumont	80.3%
Dallas	66.3%
Houston	68.4%
San Antonio	76.3%
Tyler	77.3%

7-year training period. While the former assumption is mostly true, the latter is perhaps not accurate. In Table 7, the spread of the evaluation period is much smaller than the spread of any of the training data years. When comparing 2016 to the training sample, its spread has a z -score of 3.6, so it is clear that 2016 was anomalous. This behavior does not explain our test period error, but it does highlight a limitation of this type of forecasting.

Finally, one must consider the complexity of the model when contemplating what the test error is relative to the training error. An overly complex model with respect to degrees of freedom (DOF) will yield very low error in the

training period while exhibiting substantial error in the testing period; this is because the probability of over-fitting the data increases with the DOF. Though Table 2 indicates only 8 fitted parameters, each one of the predictors has a corresponding value for the effective degrees of freedom (EDOF), which is calculated in the cubic spline of the associated smoothed function for the predictor (these smoothed functions are fitted in the GAM; see Fig. 2). For example, in the Austin quantitative model, there are 32.017 EDOF when considering the smoothing of all of the predictors. Test error as a function of degrees of freedom can be characterized as high bias and low variance at small DOF and low bias and high variance at large DOF (Hastie *et al.*, 2017). The function is minimized at some DOF, and it is possible that our models are beyond this test error minimization point.

To further assess the quality of the quantitative models, we compared the performance of the quantitative statistical model with the WRF/CMAQ-based NAQFC numerical forecast for our six urban areas. NOAA provided us with the output from NAQFS numerical forecasts for the six Texas urban areas but only for a subset of the evaluation period (19 May 2016 through 22 September 2016) used for the statistics in Table 6. Evaluation statistics for these WRF/CMAQ

Table 6. Best-fit line slope (a) and intercept (b), and Pearson correlation coefficient (r) of the $O_{3,MDA8}$ forecasted by the quantitative models as a function of the $O_{3,MDA8}$ calculated from the TAMIS O_3 measurements. The mean (μ) and standard deviation (σ) of the model-data residuals are also provided.

Urban Area	a	b	r	μ (ppbv)	σ (ppbv)
ARR	0.70	11.70	0.81	−1.04	5.74
BPA	0.65	11.98	0.77	−2.14	7.80
DFW	0.70	16.74	0.80	0.90	7.54
HGB	0.53	19.59	0.69	−4.71	9.98
SA	0.74	12.93	0.80	2.04	6.42
TLM	0.64	17.12	0.77	2.27	7.02

Table 7. The mean (μ) and standard deviation (σ) of the observed O_3 MDA8 distributions for the entire Austin training data set, each of the individual years in the training data set, and the evaluation data set (2016).

Year	$O_{3,MDA8}$ μ (ppb)	$O_{3,MDA8}$ σ (ppb)
2009–2015	46.47	13.56
2009	45.09	13.17
2010	43.45	14.65
2011	52.16	12.4
2012	46.48	13.27
2013	44.81	13.00
2014	45.64	12.01
2015	47.65	14.62
2016	43.17	9.68

forecasts are given in Table 8 (labeled “NOAA”) along with the statistics for the quantitative statistical forecasts (labeled “GAM”) in the same time period (19 May 2016 through 22 September 2016). Our statistical models have a consistently higher correlation coefficient with the measured maximum $O_{3,MDA8}$ than the NOAA numerical forecasts, and the standard deviation of the residuals is much smaller in our statistical models, both of which suggest that our statistical models are able to capture some of the variability in $O_{3,MDA8}$ that the NOAA numerical forecasts do not. Furthermore, our forecasts show a smaller absolute mean bias for all urban areas except Dallas/Fort Worth. This may be because the NOAA forecasts use the CMAQ model, which is most commonly used to investigate high O_3 events in similar large urban areas that are in non-attainment of the National Ambient Air Quality Standards (NAAQS), and thus may tend to over-predict O_3 in less polluted environments.

Probabilistic Forecasts

To assess our probabilistic model performance, we attempted to use the reliability and relative operating characteristic (ROC) area statistics, which assess how well the predicted probabilities of an event correspond to their observed frequencies and how well the model discriminates between events and non-events, respectively. For the purpose of our study, an event is defined as the $O_{3,MDA8}$ exceeding a given threshold for a given urban area on a given day. The reliability and ROC area for the models are given in Supplementary Table S4. Perfect scores for reliability are zero and for ROC area are unity, so our models ostensibly exhibit substantial capability in their event forecasting. However, it

is difficult to ascertain anything conclusive given the small number of extreme events that occurred during the evaluation period. Our test period consists of a modest sample of 184 days, but for probabilistic validation, this sample is broken down even further—first into five probability bins for both statistics (where most of the forecasts are contained in the first bin) and subsequently into events and non-events (most of which are non-events) in the ROC metric. The reliability and ROC area are thus calculated at small sample sizes. Additionally, given the R^2 in the fits to the training data (Supplementary Table S2), there is little reason to believe that the probabilistic models will be as successful as implied by Supplementary Table S4 when applied to a larger evaluation sample.

O_3 Classification Forecasts

Evaluation of classification models is routinely done using the confusion matrix utilized to assess the performance of the training set. In our study, diagonal elements of the matrix are considered successful classifications or true positives, elements above the diagonal are false negatives (i.e., a poor air quality event was missed or the severity was under-predicted by the forecaster), and elements below the diagonal are false positives (i.e., the forecaster incorrectly predicted poor air quality). The confusion matrix for each model (i.e., each urban area) is presented in Supplementary Tables S5–S10. Total model success rate is defined as the ratio of true positives to all events, which is effectively the sum of the diagonal matrices divided by the total number of elements in the confusion matrices. Likewise, the false alarm rate is defined as the ratio of false positives to all events, and the miss rate is defined as the ratio of false negatives to all events. These rates are given in Table 9.

Overall, the O_3 classification models perform well, but the tables in the supplemental data show that the more extreme the events become, the less accurate the forecasting classifier is—most of the success is due to the “green” days, and the areas that do have more “orange” days (Dallas and Houston) are significantly less accurate. This again is likely due to the lack of extreme events on which to train the models and is the random forest equivalent of the underestimation of high $O_{3,MDA8}$ days that the quantitative forecasts exhibit. The classification models can thus be considered binary air quality forecasters—they can successfully predict either “Good” or “Bad” air quality days based on EPA AQI and $O_{3,MDA8}$ specifications (https://www.tceq.texas.gov/cgi-bin/compliance/monops/ozone_summary.pl#interpret).

Table 8. Best-fit line slope (a) and intercept (b), and Pearson correlation coefficient (r) of the $O_{3,MDA8}$ forecasted by both the NOAA numerical models (i.e., the WRF/CMAQ National Air Quality Forecast Capability results for our six urban areas, labeled “NOAA” in the table) and the quantitative statistical models (labeled “GAM” in the table) as a function of the $O_{3,MDA8}$ calculated from the TAMIS O_3 measurements. The mean (μ) and standard deviation (σ) of the model-data residuals is also provided. Note: the time period for these statistics is 19 May 2016 through 22 Sep 2016.

	Urban Area	a	b	r	μ (ppbv)	σ (ppbv)
NOAA	ARR	0.42	30.62	0.32	6.04	7.71
	BPA	0.55	28.36	0.34	11.35	10.3
	DFW	0.36	35.47	0.38	0.62	10.3
	HGB	0.31	40.00	0.19	5.86	12.4
	SA	0.38	33.25	0.34	8.15	7.95
	TLM	0.58	28.64	0.42	11.69	9.14
GAM	ARR	0.62	14.12	0.79	−1.59	5.34
	BPA	0.59	13.72	0.72	−1.82	7.91
	DFW	0.70	17.49	0.80	1.14	7.80
	HGB	0.44	23.09	0.64	−4.48	10.23
	SA	0.72	12.67	0.79	1.59	5.84
	TLM	0.65	16.86	0.80	2.56	6.74

Table 9. Classification forecast success, false alarm, and miss rates for all models.

Urban Area	Success Rate	False Alarm Rate	Miss Rate
Austin	87.01%	4.54%	8.44%
Beaumont	86.36%	3.25%	10.39%
Dallas	77.27%	7.14%	15.58%
Houston	66.88%	1.95%	31.17%
San Antonio	89.12%	2.72%	8.16%
Tyler	90.26%	2.60%	7.14%

Alternative Predictors

Our forecasting scheme predicts the ozone for a given day, based on meteorological forecasts for the day and ozone measurements the day before the models are run. However, for forecasting purposes, it may be more practical to use an earlier ozone measurement or predict levels on a more intermediate scale, since $O_{3,MDA8}$ calculations are not known until shortly before our forecast. To address this issue, we replaced the $O_{3,MDA8}$ from day D_{i-1} in our models with the $O_{3,MDA8}$ from day D_{i-2} , and we predicted the $O_{3,MDA8}$ for both day D_i and day D_{i+1} .

Ozone levels are dependent on photochemistry, so it is also important to investigate the significance of solar radiation in our predictions. We accomplished this by utilizing the “CLD” field in the NAM MOS forecasts, which are classifications of total sky cover for a given hour. The categories are given in Supplementary Table S11. In training and forecasting, we assign a numerical value to the categories as shown in Supplementary Table S11 and then average over the same time period that was used for the W_S , W_D , and ρ_{WV} daily averages.

One limitation that exists with these new parameters is that the MOS forecasts only extend 6–72 hours from the NAM MOS runtime. This forecast range inhibits any of our models from predicting past day D_{i+2} . Additionally, the training data that we collected do not have forecasts for 21:00 UTC on day D_{i+2} , and this time is required for T_{2100} and the predictors that are daily averages. Consequently, our new forecasts only use the new predictors for a day D_i and

D_{i+1} forecast.

As we did for our quantitative models, we compared the predictions from these three new models with the observations by computing the linear coefficients of the best-fit lines, linear correlation coefficient, and mean and standard deviation of the differences for each model and observation set (i.e., for each of the 3 new models and 6 cities). Results are given in Supplementary Tables 12–14. When compared with Table 6, it is clear that using older ozone observations in the models and attempting to predict the levels further out is less precise than the baseline model using the predictors given in Table 2. Supplementary Table S12 provides statistics for a model where the clouds predictor is added to the baseline model and the $O_{3,MDA8}$ predictor was modified to be for day D_{i-2} . The altered model yields smaller slopes and correlation coefficients but larger best-fit intercepts, biases, and spread compared to the baseline. A model with the same predictors (clouds and $O_{3,MDA8}$ for day D_{i-2}) but whose target is the $O_{3,MDA8}$ for day D_{i+1} was also developed, and its agreement with the data exhibits similar trends as the previous model. This agreement is provided in Supplementary Table S13, which illustrates that the statistics further degrade as the prediction becomes more intermediate rather than short-term.

One more model we trained was the baseline model (Table 2) with only the clouds predictor added to isolate the effect of including cloud forecasts. Statistics are provided in Supplementary Table S14. Again, aggregate agreement worsens but only slightly compared to the baseline. It is worth

noting that the slope and bias for San Antonio improve, as does the bias for Dallas, though for both of these sites, the spread in the model-data differences increases.

CONCLUSIONS AND FUTURE WORK

For six urban areas in eastern Texas, we developed three types of forecasting models that predict the maximum daily 8-hour-averaged O_3 ($O_{3,MDA8}$) for a given day based on meteorological forecasts for the same day and the $O_{3,MDA8}$ from the previous day. The three types of modeling were quantitative forecasting of $O_{3,MDA8}$ (based on a GAM with a log-link function), probabilistic forecasting that predicted the probability of $O_{3,MDA8}$ being above a given threshold (based on a GAM with a logit function), and random forest classification for predicting the AQI classifications. We performed a number of sensitivity studies to find the best set of predictors to use for all three types (and for all six urban areas). These parameters included the MOS-forecasted afternoon high temperature; MOS-forecasted daily temperature difference; daily averaged MOS wind speed, wind direction, and water vapor density; day of the week; day of the year; and $O_{3,MDA8}$. The models were then tested by forecasting the 2016 ozone season (May through October).

The quantitative forecasts explain 69% or more of the variance in each urban area except for TLM. Most of the predictors were statistically significant, and those there were not varied with location. The probabilistic forecasting models exhibited little accuracy and thus are not recommended for forecasting unless significant improvements can be implemented. The ozone classification forecast using the random forest technique correctly predicted the O_3 AQI classification 67% of the time in HGB, 77% of the time in DFW, and over 86% of the time elsewhere. False negatives (where the severity of a poor air quality event is under-predicted) were generally more common than false positives (where air quality is predicted to be worse than observed).

While the quantitative and classification forecasting models exhibit accuracy in forecasting O_3 and can thus be used to assist with issuing public statements and advisories regarding air quality, issues remain that must be addressed in future work. The quantitative model forecasts trend toward the mean and have less variance than the observations, resulting in slopes for the lines best-fitting the model and the data that fall between 0.5 and 0.75 rather than near unity. Additionally, our classification scheme can best be described as binary because it correctly identifies good (green) and poor air quality days but fails to distinguish between different degrees of elevated O_3 (yellow, orange, red, and maroon). The reason for this characteristic is likely the small number of extreme events in the training set.

Funding limited the scope of this project to urban areas in Texas, but future research will focus on expanding the models to work in areas outside of Texas and improving their forecasting performance. Extending the forecasts to other states and perhaps countries should be possible, but more work is needed to determine how the methods in this paper will perform outside of Texas. We intend to address the performance with techniques such as weighting the

mentioned extreme events more heavily in the model training and applying neural networks instead of generalized additive models.

ACKNOWLEDGMENTS

This study was funded by TCEQ Contract No. 582-15-50414 to AER, and we would also like to thank E. Gribbin of TCEQ for providing meteorological and air quality data for our training and forecast evaluation. We thank Pius Lee, Jeff McQueen, and Daniel Tong of NOAA for providing us with their numerical O_3 forecasts that we used in our validation. Finally, we thank the UCAR Data Archive support staff, specifically Douglas Schuster and Chi-Fan Shih, for greatly facilitating the collection of historical NCEP NAM-12 model output. Lastly, we express our gratitude to the anonymous reviewers that enriched the content of this publication.

SUPPLEMENTARY MATERIAL

Supplementary data associated with this article can be found in the online version at <http://www.aaqr.org>.

REFERENCES

- Abdul-Wahab, S.A. and Al-Alawi, S.M. (2002). Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. *Environ. Modell. Software* 17: 219–228.
- Anderson, H.R., Spix, C., Medina, S., Schouten, J.P., Castellsague, J., Rossi, G., Zmirou, D., Touloumi, G., Wojtyniak, B., Ponka, A., Bacharova, L., Schwartz, J. and Katsouyanni, K. (1997). Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: Results from the APHEA project. *Eur. Respir. J.* 10: 1064–1071.
- Bell, M., McDermott, A., Zeger, S., Samet, J. and Dominici, F. (2004). Ozone and short-term mortality in 95 US urban communities, 1987–2000. *JAMA* 292: 2372–2378.
- Bey, I., Jacob, D.J., Yantosca, R.M., Logan, J.A., Field, B.D., Fiore, A.M., Li, Q., Liu, H.Y., Mickley, L.J. and Schultz, M.G. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res.* 106: 23073–23095.
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45: 5–32.
- Brown, J., Bateson, T. and McDonnell, W. (2008). Effects of exposure to 0.06 ppm ozone on FEV1 in humans: A secondary analysis of existing data. *Environ. Health Perspect.* 116: 1023–1026.
- Burnett, R., Brook, J., Yung, W., Dales, R. and Krewski, D. (1997). Association between ozone and hospitalization for respiratory diseases in 16 Canadian cities. *Environ. Res.* 72: 24–31.
- Byun, D. and Schere, K.L. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) system. *Appl. Mech. Rev.* 59: 51–77.

- Camalier, L., Cox, W. and Dolwick, P. (2007). The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmos. Environ.* 41: 7127–7137.
- Carlton, A.G., Bhawe, P.V., Napelenok, S.L., Edney, E.O., Sarwar, G., Pinder, R.W., Pouliot, G.A. and Houyoux, M. (2010). Model representation of secondary organic aerosol in CMAQv4.7. *Environ. Sci. Technol.* 44: 8553–8560.
- Carter, G., Dallavalle, J. and Glahn, H. (1989). Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Weather Forecasting* 4: 401–412.
- Devlin, R., Raub, J. and Folinsbee, L. (1997). Health effects of ozone. *Sci. Med.* 4: 8–17.
- Feister, U. and Balzer, K. (1991). Surface ozone and meteorological predictors on a subregional scale. *Atmos. Environ.* 25: 1781–1970.
- Glahn, B., Gilbert, K., Cosgrove, R., Ruth, D. and Sheets, K. (2008). The gridding of MOS. *Weather Forecasting* 24: 520–529.
- Gong, X., Kaulfus, A., Nair, U. and Jaffe, D. (2017). Quantifying O₃ impacts in urban areas due to wildfires using a generalized additive model. *Environ. Sci. Technol.* 51: 13216–13223.
- Janjic, Z.I. (2003). A nonhydrostatic model based on a new approach. *Meteorol. Atmos. Phys.* 82: 271–285.
- Kgabi, N.A. and Sehloho, R.M. (2012). Tropospheric ozone concentrations and meteorological parameters. *Global J. Sci. Front. Res.* 12: 11–21.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Software* 28: 1–26.
- Lippmann, M. (1989). Health effects of ozone: A critical review. *JAPCA* 39: 672–695.
- Logan, J. (1985). Tropospheric ozone: Seasonal Behavior, Trends, and Anthropogenic Influence. *J. Geophys. Res.* 90: 10463–10482.
- Luna, A.S., Paredes, M.L., de Oliveira, G.C. and Correa, S.M. (2014). Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. *Atmos. Environ.* 98: 98–104.
- National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, European Centre for Medium-Range Weather Forecasts, and Unidata/University Corporation for Atmospheric Research. (2016, April 26). *Historical Unidata Internet Data Distribution (IDD) Gridded Model Data. 2003, updated daily*. Retrieved from Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory: <http://rda.ucar.edu/datasets/ds335.0/>
- Pan, L., Tong, D., Lee, P., Kim, H.C. and Chai, T. (2014). Assessment of NO_x and O₃ forecasting performances in the U.S. National Air Quality Forecasting Capability before and after the 2012 major emissions updates. *Atmos. Environ.* 95: 610–619.
- Rutledge, G., Alpert, J. and Ebisuzaki, W. (2006). NOMADS: A climate and weather model archive at the national oceanic and atmospheric administration. *Bull. Am. Meteorol. Soc.* 97: 327–341.
- Thompson, M., Reynolds, J., Cox, L.H., Guttorp, P. and Sampson, P.D. (2001). A review of statistical methods for meteorological adjustment of tropospheric ozone. *Atmos. Environ.* 35: 617–630.
- U.S. Environmental Protection Agency (2013). *U.S. EPA. Integrated Science Assessment (ISA) of Ozone and Related Photochemical Oxidants*. Washington, DC, DC, USA, EPA/600/R-10/076F.
- Wood, S.N. (2006). *Generalized additive models: An introduction With R*. Chapman and Hall/CRC, Boca Raton, FL.
- Yarwood, G. (2005). Updates to the carbon bond chemical mechanism: CB05. http://www.camx.com/publ/pdfs/cb05_final_report_120805.pdf
- Yu, R., Yang, Y., Yang, L., Han, G. and Move, O.A. (2016). RAQ—A random forest approach for predicting air quality in urban sensing systems. *Sensors* 16: 86.

Received for review, December 12, 2018

Revised, June 26, 2019

Accepted, October 6, 2019