Atmospheric Environment 112 (2015) 1–11

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Climate controls on air quality in the Northeastern U.S.: An examination of summertime ozone statistics during 1993–2012

Evan M. Oswald [a, *], Lesley-Ann Dupigny-Giroux [b], Eric M. Leibensperger [c], Rich Poirot [d], Jeff Merrell [d]

[a] Postdocs Applying Climate Expertise Fellowship Program, University Corporation for Atmospheric Research, Boulder, CO, USA
[b] Department of Geography, University of Vermont, Burlington, VT, USA
[c] Center for Earth and Environmental Science, SUNY Plattsburgh, Plattsburgh, NY, USA
[d] Air Quality and Climate Division, Vermont Department of Environmental Conservation, Vermont Agency of Natural Resources, Montpelier, VT, USA

HIGHLIGHTS

- We model surface ozone concentrations using meteorological and climate variables.
- Meteorological-driven statistical models could improve ozone season predictions.
- Teleconnection-driven statistical models were only insightful.
- Precipitation, temperature and solar radiation were strong predictors.
- Pacific Decadal, Quasi-Biennial and Arctic Oscillations were superior predictors.

ARTICLE INFO

Article history:
Received 23 October 2014
Received in revised form
8 April 2015
Accepted 9 April 2015
Available online 11 April 2015

Keywords:
Tropospheric ozone
Meteorology
Teleconnections
Northeastern U.S.

ABSTRACT

The goal of this study is to better understand the linkages between the climate system and surface-level ozone concentrations in the Northeastern U.S. We focus on the regularity of observed high ozone concentrations between May 15 and August 30 during the 1993–2012 period. The first portion of this study establishes relationships between ozone and meteorological predictors. The second examines the linkages between ozone and large-scale teleconnections within the climate system. Statistical models for each station are constructed using a combination of Correlation Analysis, Principal Components Analysis and Multiple Linear Regression. In general, the strongest meteorological predictors of ozone are the frequency of high temperatures and precipitation and the amount of solar radiation flux. Statistical models of meteorological variables explain about 60–75% of the variability in the annual ozone time series, and have typical error-to-variability ratios of 0.50–0.65. Teleconnection patterns such as the Arctic Oscillation, Quasi-Biennial Oscillation and Pacific Decadal Oscillation are best linked to ozone in the region. Statistical models of these patterns explain 40–60% of the variability in the ozone annual time series, and have a typical error-to-variability ratio of 0.60–0.75.

1. Introduction

Tropospheric ozone is primarily formed in the atmosphere as the result of photochemical reactions between precursor species of nitrogen oxides (NO_x) and volatile organic compounds (VOCs). Meteorology plays a major role by influencing chemical reaction rates of ozone formation and destruction; emission rates of VOC and NO_x precursors; as well as atmospheric mixing, the accumulation and transport of ozone and precursors to downwind receptor locations. Establishing the controls on ozone is difficult since long-term ozone observations are sparse and uncertainties associated with emissions and meteorological conditions can be substantial. This constrains air quality professionals in their need to provide both short-term forecasts and long-term planning.

Early ozone control strategies focused primarily on reducing urban VOC emissions. Increasing recognition of the importance of ozone transport led to a concerted effort to reduce NO_x emissions – for example as recommended by the multi-state Ozone Transport Assessment Group after years of modeling and data analysis (LeClair, 1997). Such work led the U.S. Environmental Protection

* Corresponding author. Geography Department, University of Vermont, 94 University Place, Burlington, VT 05405-0114, USA.
E-mail address: eoswald@uvm.edu (E.M. Oswald).

http://dx.doi.org/10.1016/j.atmosenv.2015.04.019

Agency (EPA) to enact the "NO_x SIP Call" in 1998, requiring reduced emissions in 22 states by 2003. The general consensus is that anthropogenic NO_x emissions reductions are the best way to reduce surface ozone, and that modern NO_x reductions have driven a substantial decline in surface ozone concentrations (Baldridge, 2005).

Daily maximum air temperatures correlate well with ozone concentrations in the Eastern (Rao et al., 2003; Rasmussen et al., 2012) and Northeastern U.S. (Bloomer et al., 2009). Of all the individual meteorological variables, it is generally believed that temperature has the strongest relationship with ozone (Jacob and Winner, 2009). Other meteorological variables related to ozone include wind speed (Vukovich, 1994; Rao et al., 2003; Chan, 2009), mixing layer height (Rao et al., 2003), boundary layer ventilation (Rao et al., 2003), surface pressure (Vukovich, 1994) and surface radiation flux (Vukovich, 1994; Rao et al., 2003). Atmospheric moisture content is another meteorological variable linked to surface-level ozone (Chan, 2009; Rao et al., 2003), as are frontal passages (Leibensperger et al., 2008) and stagnation events (Vukovich, 1995).

In the Northeastern U.S., ozone is also dependent upon regional transport, making the wind direction an integral factor in ozone concentrations (Husar and Renard, 1997; Schichtel and Husar, 2001). Unlike most regions where low wind speeds are associated with high ozone levels, higher wind speed typically aids in effective transport and is positively correlated with ozone levels (Husar and Renard, 1997; Schichtel and Husar, 2001).

It is well known that modes of climate system variability, or "teleconnections", across the Northern Hemisphere play a significant role in climate changes on interannual to interdecadal time scales. Thus, when investigating changes in tropospheric ozone, teleconnections should also be considered (Lin et al., 2014). Recent work has linked teleconnections and surface-level ozone (e.g. Lin et al., 2015; Lin et al., 2014). Here an exploratory approach was taken and the physical mechanisms were discussed only for teleconnection patterns found to have statistical relationships with ozone.

This study contributes to existing literature on the influence of weather and climate on ozone by examining:

- the upper tail of the ozone distribution via the 60 ppbv threshold
- a large suite of meteorological variables, previously unanalyzed simultaneously
- meteorological variables not previously examined
- the physical mechanisms underlying the meteorology-ozone relationship, thereby moving beyond previous work on the removal of the influence of meteorology (e.g. Milanchus et al., 1998)
- the linkages between ozone and hemispheric teleconnections

In this study, research questions focus on the interplay among ozone, meteorological variables and teleconnections. Can empirical models of meteorological predictors and NO_x emissions explain ozone changes over time? What meteorological predictors drive the majority of these models? Can similar models built on teleconnections and NO_x emissions explain ozone changes over time? Which teleconnections are important predictors?

2. Material and methods

2.1. Focus region

The Northeastern U.S. region is downwind of the majority of the continental U.S. and marks the exit region of the Polar Jet Stream. The region is topographically complex, extensively borders the Atlantic Ocean and frequently features a low level jet. The general pattern of NO_x emissions in the region decreases from higher to lower values moving in the northeast direction.

2.2. Ozone data

Ozone observations were acquired from the EPA's Air Quality System. Tools on the Datafed.net website (Husar et al., 2008) facilitated the spatial delineation and extraction of the 1993–2012 time series of daily maximum 8-hour average values of ozone (MDA8O3) for all available (253) stations in the Northeastern U.S. region (39.4°-45.4°N, 69.5°-80.2°W). To minimize the influence of missing data, each station used was required to have at least 80% of the years during the 1993–2012 period with 80% of the dates in each year's "season" (defined as May 15-August 30). We chose this period over the April 1-September 30 period because the smaller window generally had higher ozone levels. Stations meeting our completeness criteria constituted the set of stations (83) used in this study (Fig. 1).

Two metrics were derived from the ozone observations. The primary metric to quantify ozone for this analysis was the percent of the season with the MDA8O3 $\geq$60 ppbv. This will be referred to as the "60 ppbv" metric. While the 60 ppbv level is well below the level of the current (2008) U.S. National Ambient Air Quality Standard (NAAQS) of 75 ppbv, it is within the range recommended by the EPA's Clean Air Scientific Advisory Committee during the last two rounds of ozone NAAQS review (e.g. Frey, 2014). The second metric was the MDA8O3 value corresponding to the 80th percentile over the season and is referred to as "80 pctl". The 60 ppbv concentration level was chosen to better reflect human and ecological health-relevant exposures, and the 80th percentile correlated with the 60 ppbv metric better than other percentile levels. Using two ozone metrics instead of one allowed for more concrete conclusions about the relationships between high ozone concentrations and predictor metrics.

2.3. Meteorological and climate data

Gridded meteorological data in this study were from one of three climate data products, depending upon the variable and geographical location. Within the U.S., daily total precipitation, daily maximum and minimum temperatures were extracted from the 4 km resolution PRISM AN81d dataset (Daly et al., 2008). The same variables were extracted from the DAYMET dataset (Thornton et al., 1997) for the Canadian provinces. DAYMET is a 1 km resolution daily dataset that was aggregated to 4 km resolution and regridded to that of the PRISM dataset for use in this study.

The North American Regional Reanalysis (NARR) (Mesinger et al., 2006) provided the other meteorological variables, including daily average relative humidity, 40–100 cm volumetric soil moisture, downwelling shortwave radiation flux, 10m above ground wind speed and direction, 850 mb geostrophic wind speed and direction, mean sea level pressure and 500 mb geostrophic wind speed. NARR data consist of daily values on a 30 km resolution grid, which were regridded to the PRISM grid via inverse distance weighting.

The teleconnections indices were preferentially acquired from the NOAA's Physical Science Division <http://www.esrl.noaa.gov/psd/data/climateindices/list/>, followed by the National Weather Service's (NWS) Climate Prediction Center <http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml>.

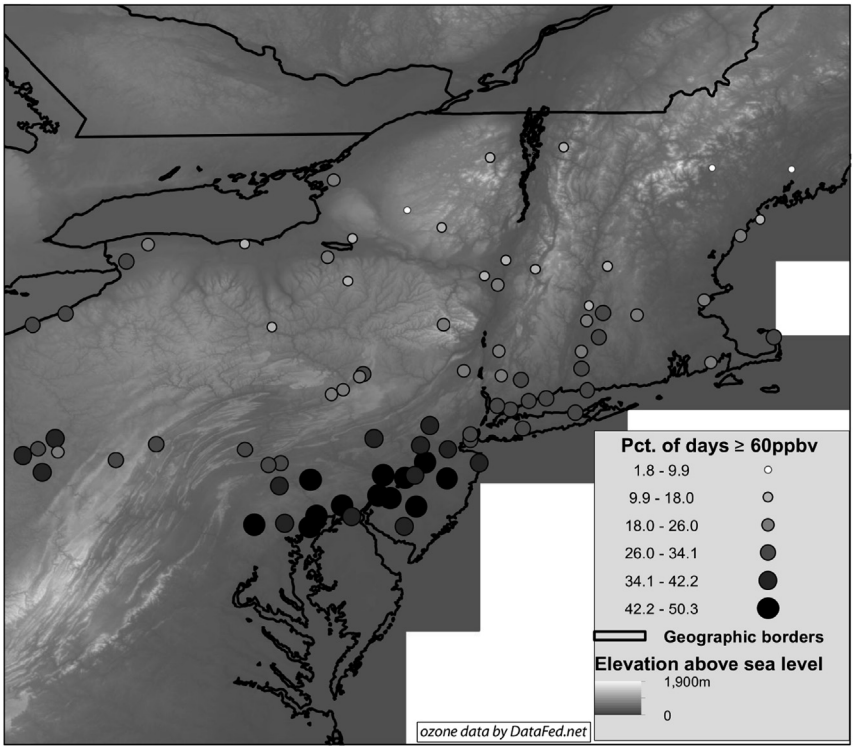

Fig. 1. Map of the 83 stations in the Northeastern U.S. showing the spatial distribution of the 20-year mean (1993–2012) percent of days with MDA8O3 ≥60 ppbv during 15 May – 30 August. Symbology by equal interval groupings. Elevation data provided by USGS.

2.4. Predictor metrics

Predictor metrics belong to one of four "types" (surface climate, weather, air-flow and teleconnections). The first three types together were considered "meteorological" predictor metrics, with unique values for each station extracted from the meteorological grids at the locations of the ozone stations. The teleconnections type metrics were considered separately and had the same values regardless of the station.

2.4.1. Surface climate type metrics

The 20 selected surface climate type metrics focused on temperature, solar radiation, atmospheric moisture, precipitation and soil moisture (Table 1). Metrics of the first three foci have been linked to ozone in the literature (as cited in Section 1): temperature through reaction rates (decomposition of peroxyacyl nitrates) and VOC emissions; solar radiation flux through photochemical reaction rates and; atmospheric moisture through chemical reaction rates (removal via hydroxyl radical). Precipitation has not been formally linked to ozone, but could be influential through atmospheric moisture content or the scavenging of nitrogen species (Crutzen and Lawrence, 2000). Soil moisture metrics were also not previously linked with ozone, but process level mechanisms hint at a possible indirect relationship through temperature and atmospheric moisture (Fischer and Seneviratne, 2007). Surface climate type metrics come in local and regional varieties. For each ozone station used, local metrics were extracted from the grid cell encompassing it, while the regional metrics were extracted from all grid cells with centroids within 111 km (i.e. 1° latitude) of the station. This distance represented a relevant increase in spatial extent from the local scale. Percentiles were calculated based on the 1981–2010 period, and were specific to both the calendar date and geographic location. Local soil moisture could not be calculated at 13 coastal stations due to the absence of soil moisture values at these locations in the NARR dataset. Regional soil moisture values were computed for all 83 stations.

Table 1
Surface climate-type metrics used in this study.

Predictor metric
Mean daily maximum temperature
Number of days with daily maximum temperature ≥ the 90th percentile
Season average daily mean dewpoint temperature
Number of days with dewpoint temperature ≥ the 90th percentile
Number of days with ≥0.01″ ("trace") of precipitation
Daily mean volumetric soil moisture
Number of days with soil moisture ≥ 90th percentile
Number of days with soil moisture ≤ 10th percentile
Mean daily total downwelling shortwave radiation flux
Number of days with downwelling shortwave radiation flux ≥ 66th percentile

2.4.2. Weather type metrics

Some metrics, e.g. the number of stagnation events at a given station, were better classified as "weather" rather than surface climate. Stagnation metrics were separated into the number of events and total days per season. We followed the Wang and Angell (1999) definition: stagnation events as three days of mean sea level geostrophic wind speeds < 8 m/s, with no precipitation and 500 mb geostrophic wind speed <13 m/s. Physically, stagnation is linked to ozone concentrations through the trapping of ozone and ozone precursor species (Valente et al., 2012).

The second weather metric used was the number of frontal passages at each station. Frontal passages are physically related to ozone due to their ability to ventilate ozone and its precursors from a location (Leibensperger et al., 2008). Frontal passages were diagnosed using time series of three variables: maximum temperature, dewpoint temperature and wind direction. For each variable

at a given station, the mean and standard deviation (over all years and dates) of the absolute difference between one day and the preceding day were calculated at each station and each variable. A difference or step function larger than one standard deviation above the mean was designated as a "jump". When two or more variables had a simultaneous jump, a frontal passage was identified. The typical number of frontal passages (7.5 per season) and stagnation events (1.8 per season) agreed with other studies (Wang and Angell, 1999; Leibensperger et al., 2008). Our diagnosis of frontal passages also compared well with historical maps from the online Weather Prediction Center Surface Analysis Archive (at <http://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive.php>, not shown).

2.4.3. Air-flow metrics

Using the observed relationships between wind characteristics and ozone, four air-flow metrics were created to represent regional transport at each station. These metrics were based on previously demonstrated differentiation of ozone concentrations in the Northeastern U.S. by local wind speed and direction (Husar and Rendard, 1997). Thus, each date was classified into 18 different "bins" based on both the wind speed (low, medium, high) and direction (60-degree bins). The mean of each bin was calculated, as well as the percent of the bin sample that was $\geq$60 ppbv.

Two kinds of air-flow metrics were computed, "simple" and "sophisticated". The former was a count of the number of dates in a given season that fell into one of the top 4 bins of the bin-mean ozone values. The latter was the product of the number of instances that each bin occurred during a season and the percent of the time that the bin was $\geq$60 ppbv. Metrics were also calculated at both a height of 10-meters above ground level as well as the geostrophic winds at 850 mb.

2.4.4. Teleconnection metrics

Sixteen metrics based on teleconnections were computed from the seasonal means of the indices representing the Arctic Oscillation (AO), North Atlantic Oscillation (NAO), Atlantic Multidecadal Oscillation (AMO), Pacific Decadal Oscillation (PDO), Quasi-Biennial Oscillation (QBO) and the Tropical/Northern Hemisphere pattern (T/NH). Only the seasonal means that coincided with the teleconnection patterns being a leading mode of variability were included. The September-October-November (SON) means were excluded because preliminary examination indicated that they were never correlated with the ozone metrics. The Earth System Research Laboratory Physical Sciences Division's "Linear Correlations in Atmospheric Seasonal/Monthly Averages" tool <http://www.esrl.noaa.gov/psd/data/correlation/> was used to diagnose the statistical associations among the selected teleconnections and temperature and precipitation.

The AMO index describes the pattern of North Atlantic sea surface temperature variability, which is correlated to temperature and precipitation across much of the Northern Hemisphere, as well as with hurricanes in the Atlantic Ocean (Enfield et al., 2001). The inter-related AO and NAO indices describe the circumpolar vortex strength, which in a general sense controls the likelihood of cold air outbreaks, the position of storm paths and blocking patterns in most of the northern U.S. (Thompson and Wallace, 2000). While the PDO index is best known to describe a multi-decadal oscillation, on interannual time scales, it represents the sum of ENSO-induced and random variability in the Aleutian Low. Thus, its impacts on surface climate are similar to the ENSO (Mantua et al., 1997). The QBO is an index describing a tropical stratosphere zonal-wind pattern that modulates the circumpolar stratosphere circulations by modifying the Eliassen-Palm flux, and ultimately the polar vortex strength (Baldwin et al., 2001). The T/NH index depicts a pattern of winter surface pressure anomalies in the Gulf of Alaska and Hudson Bay. In the U.S. it is commonly associated with the Pacific Polar Jet Stream positioning, the strength of the Hudson Bay Low and marine air inflow (Mo and Livezey, 1986).

2.4.5. Emissions of ozone precursor species

Temporal changes in NO_x emissions during the 1993–2012 period were quantified with the gridded emissions data that supported the CMIP5 project (Taylor et al., 2012). This ensured availability of future emissions projections that could be used with climate projections in subsequent studies focusing on climate change impacts. Past anthropogenic (excluding shipping and aviation) and burning emissions (Lamarque et al., 2010) were used along with the equivalent RCP 4.5 and RCP 8.5 emissions (Meinshausen et al., 2011). The 1993–2012 time series for May, June, July and August-mean monthly NO_x emissions (kg/m^2) were calculated over the surrounding region (25–55° N latitude and 67–110° W longitude). This compared well with the national average NO_2 concentrations time series (not shown).

2.5. Analysis

2.5.1. Emissions' role

Prior to examining the relationships between ozone and meteorology, the influences of NO_x emission changes were accounted for by linearly regressing the NO_x time series (section 2.4.5) against each station's time series of ozone (both metrics separately). Subsequent analyses used these residuals (i.e. detrended ozone time series). It was assumed the study region did not switch between NO_x- and VOC-limited regimes during the 1993–2012 period. The treatment of the NOx emission influences was simplified because this paper is not focused on NOx influences.

2.5.2. Correlation analysis

Spearman Rank correlation coefficients were calculated between the time series of predictor metrics and ozone metrics, ozone time series between different stations and between ozone metrics at the same station. Predictor metrics with absolute magnitudes of correlation (with ozone metrics) greater than or equal to the value of the 75th percentile became candidates for Principal Components Analysis (PCA) analysis. This effectively selected the 7–8 highest correlated predictor metrics for the meteorological metrics and the 3–4 highest correlated predictor metrics for the teleconnection metrics.

2.5.3. Principal Components Analysis (PCA)

PCA reduced the multi-collinearity in the set of correlation-selected predictor metrics. PCA is a statistical tool for maximizing the correlations or covariances between predictors and predictands by creating new variables ("principal components") that are both orthogonal to one another and comprised of linear combinations of the predictor metrics. The correlation matrix, instead of covariance matrix, was used in the PCA since the units of the predictor metrics widely differed one another. Varimax Rotation of the principal components ensured that each predictor metric dominated only one principal component, and illuminated the amount of variability in the set of predictors explained by each predictor variable. The dominant predictor metrics of the principal components that explained at least 5% of the variability were selected for model building. This effectively selected 2–5 metrics for the meteorological set of predictors and 2–4 for the teleconnection set.

2.5.4. Multiple Linear Regression

For each station, statistical models were built via Multiple Linear Regression (MLR) using the PCA-selected predictor metrics chosen

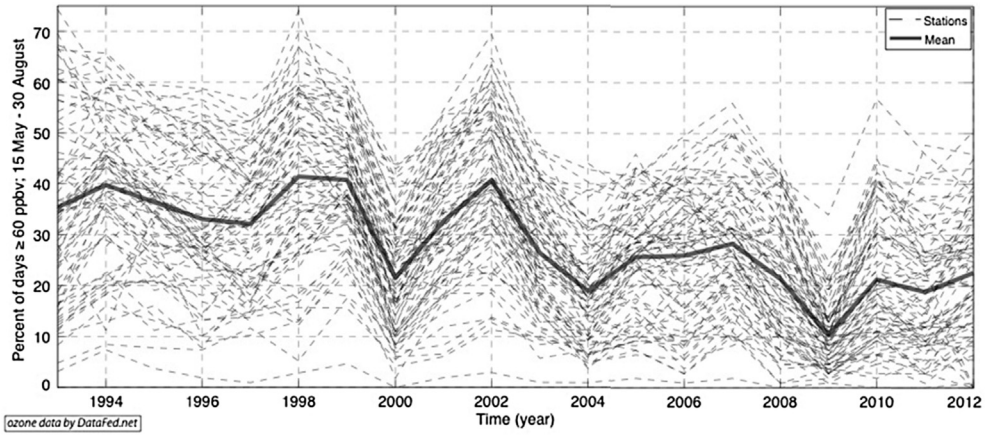

Fig. 2. Time series of the percent of days with MDA8O3 ≥60 ppbv during 15 May – 30 August of the 1993–2012 time period for all 83 study stations. The thick, solid line is the arithmetic mean of the 83 stations.

for that ozone time series. While acknowledging that this choice may have reflected hidden biases given the non-linear and non-additive nature of the underlying processes, the MLR method was straightforward, easy to interpret and the most common method of building statistical models between ozone and meteorological variables (Thompson et al., 2001). Both the predictor metrics and ozone predictands were normalized so that regression coefficients reflected the predictor metrics' relative importance.

The models needed to be reliable outside of the period upon which they are built. Therefore, the MLR was used in a "leave one out" approach, where each year in the time series was iteratively omitted and the regression coefficients estimated using the remaining years. This provided the same number of regression coefficient estimates as years in the time series. The regression coefficients were then estimated by the median value of those regression coefficients. Thus, the observations were not directly used to calculate the regression coefficients.

2.5.5. Evaluating the constructed models

Recognizing the limitations of using a 20-step time series, we exercised caution in interpreting the underlying processes extracted from the model-building methodology above. For example, prediction metrics over the entire 83-station sample were used for interpretation rather than the spatial distributions of the results that were overly noisy. Furthermore, our examination focused on the predictor metrics that were successful in both ozone metrics.

The performance of the constructed models outside of the 1993–2012 period was estimated during model building as well as in the validation phase. The "leave one out" cross-validation

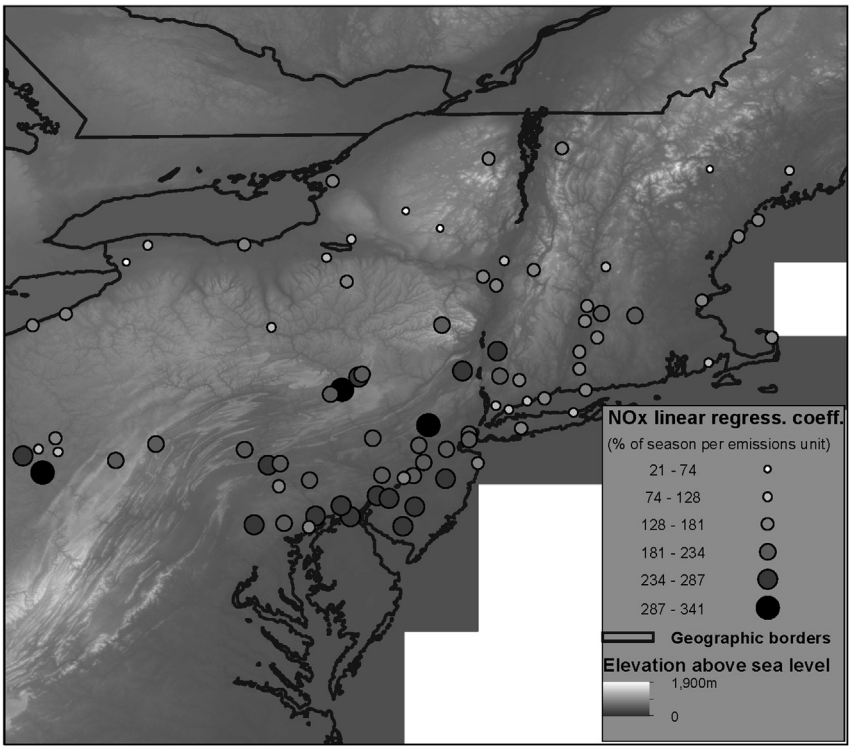

Fig. 3. Map of the 83 stations in the Northeastern U.S. showing the spatial distribution of the linear regression coefficients between regional NOx emissions and percent of days with MDA8O3 ≥60 ppbv during 15 May – 30 August. Symbology and elevation the same as shown in Fig. 2.

approach facilitated evaluation of the models during the building stage. Evaluation of the models after the building process used the model estimates and the observations. The cross validation method is typically more accurate than estimating the uncertainty directly against the training set of observations. However, since the regression coefficients were built using the training set of observations indirectly, the true accuracy of the model outside of the training period may have been closer to the performance of the models after the building process than is usually the case.

Three statistical measures quantified the performance of the models. The first was the coefficient of determination (R^2), which is the percentage of the total variance of ozone explained by the models. The next two metrics described the uncertainty in the models relative to the temporal variability in the ozone metrics. The first, the "RMSE ratio", was the ratio of the root mean squared error to the root mean squared deviation from the mean of the observations (Equation (1)), while the second, the "MAD ratio", was the ratio of the median of the absolute error in the model to the median of the absolute deviation from the median of the observations (Equation (2)). The RMSE ratio is sensitive to extremes while the MAD ratio is not.

$$\mathrm{RMSE} = \frac{\sqrt{mean\left(|x_i - y_i|^2\right)}}{\sqrt{mean\left(|x_i - mean(x)|^2\right)}} \quad (1)$$

$$\mathrm{MAD} = \frac{median(|x_i - y_i|)}{median(|x_i - median(x)|)} \quad (2)$$

where x are the ozone time series values, y are the model predicted ozone time series and i are the different years in the time series.

3. Results

3.1. Background and emissions

The two ozone metrics (60 ppbv and 80 pctl) correlated well with each another, with a mean correlation (over the 83 stations) of 0.93. With a mean correlation of 0.71 in the 60 ppbv metric and 0.73 in the 80pct metric, less agreement was seen between different stations of the same ozone metrics (Fig. 2). The distribution of the 1993–2012 mean in the 60 ppbv metrics ranged from 1.9 to 50.2 percent, with an average of 28.6 percent. The 80 pctl metric means ranged from 42.6 to 76.6 ppbv with a 64.6 ppbv average. The spatial distribution of ozone metrics' temporal means (Fig. 1) indicated a general proliferation in the south and southwest, as well as an increase with proximity to the ocean. This pattern was observed in both ozone metrics. Small-scale spatial variability was also very noticeable.

Accounting for the effects of NO_x changes produced detrended time series and regression coefficients at each station. These regression coefficients (percent of season with MDA8O3 ≥60 ppbv,

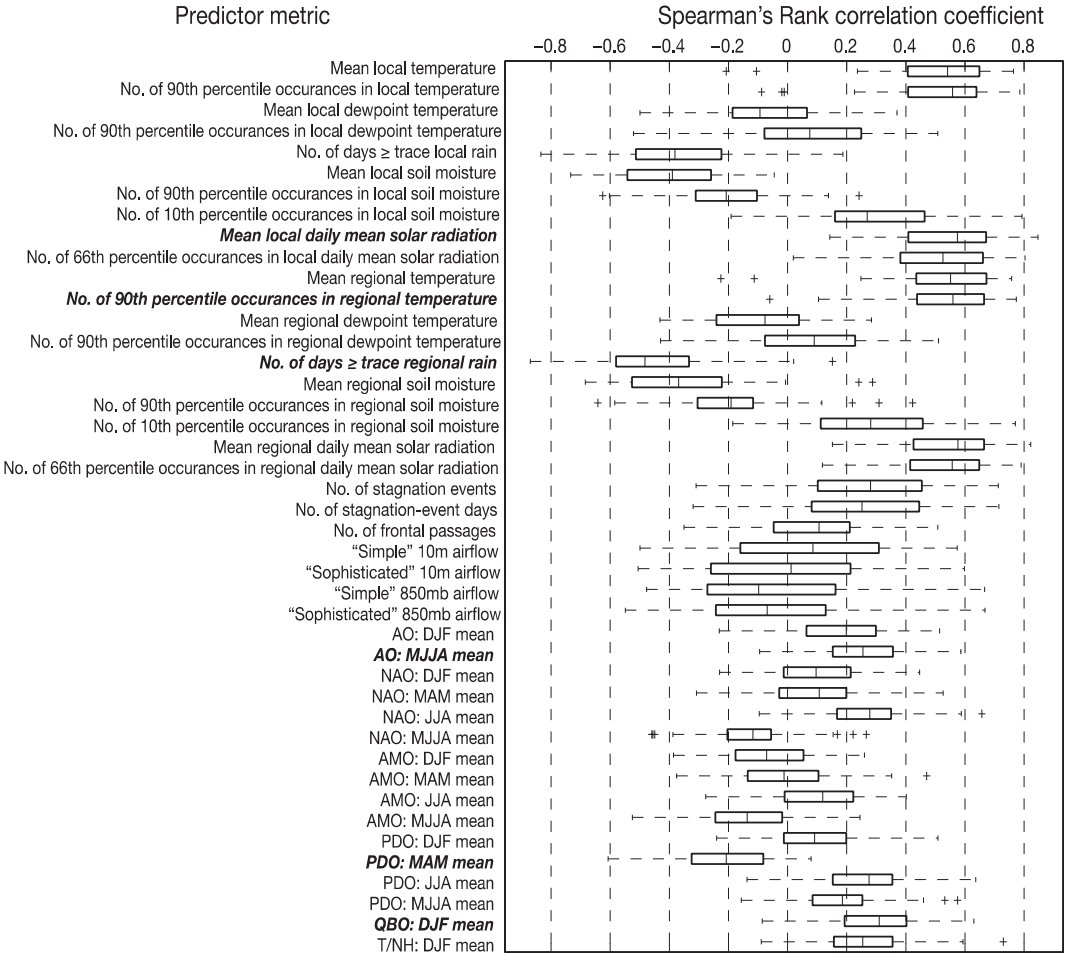

Fig. 4. Boxplots of the Spearman Rank correlation coefficients between percent of days with MDA8O3 ≥60 ppbv during 15 May – 30 August and various predictor metrics for the 83 study stations. Metrics shown in bold were those most often included in the regression models.

Table 2
Number of stations at which each meteorological predictor metric was selected for PCA selection process[a].

Predictor metric	60 ppbv	80 pctl
Mean local temperature	45	32
No. of 90th percentile occurrences in local temperature	45	57
Mean local dewpoint temperature	1	1
No. of 90th percentile occurrences in local dewpoint temperature	6	5
No. of days ≥ trace local rain	13	15
Mean local soil moisture	9	14
No. of 90th percentile occurrences in local soil moisture	0	1
No. of 10th percentile occurrences in local soil moisture	6	8
Mean local daily mean solar radiation	55	41
No. of 66th percentile occurrences in local daily mean solar radiation	47	47
Mean regional temperature	56	41
No. of 90th percentile occurrences in regional temperature	58	64
Mean regional dewpoint temperature	0	0
No. of 90th percentile occurrences in regional dewpoint temperature	8	4
No. of days ≥ trace regional rain	29	43
Mean regional soil moisture	8	23
No. of 90th percentile occurrences in regional soil moisture	3	2
No. of 10th percentile occurrences in regional soil moisture	5	5
Mean regional daily mean solar radiation	54	46
No. of 66th percentile occurrences in regional daily mean solar radiation	53	57
No. of stagnation events	13	16
No. of stagnation-event days	13	13
No. of frontal passages	2	3
“Simple” 10 m airflow	12	10
“Sophisticated” 10 m airflow	10	9
“Simple” 850 mb airflow	6	5
“Sophisticated” 850 mb airflow	9	4
AO: DJF mean	27	13
AO: MJJA mean	36	29
NAO: DJF mean	13	14
NAO: MAM mean	12	6
NAO: JJA mean	38	33
NAO: MJJA mean	12	27
AMO: DJF mean	2	1
AMO: MAM mean	6	7
AMO: JJA mean	13	7
AMO: MJJA mean	13	10
PDO: DJF mean	12	19
PDO: MAM mean	26	30
PDO: JJA mean	37	34
PDO: MJJA mean	10	19
QBO: DJF mean	45	46
T/NH: DJF mean	30	35

[a] The number of stations is out of 83 except for the local soil moisture metrics (out of 70); correlations calculated between predictor metrics and detrended ozone time series; both ozone metrics are shown; metrics in bold were those most often included in the regression models.

or MDA8O3 of the 85th percentile, per monthly NO_x emissions in kg/m^2) are measures of the linear trends that were detected and removed in the ozone time series (Fig. 3). All trends were negative. Spatial patterns of regression coefficients agreed across ozone metrics and were similar to the pattern of temporal averages (e.g. Fig. 1), which implied the locations of highest ozone levels frequently showed the most decline.

3.2. Correlation analysis

Correlations between ozone predictor metrics and the detrended ozone time series were highly variable across the 83 stations and predictor metrics, as summarized in boxplots (Fig. 4). The frequency with which each predictor metric was selected via correlation ranking (Table 2) illustrated the strength of the metric's correlations with ozone. The meteorological predictor metrics that showed strong positive correlations with the ozone metrics were predominantly temperature-based ones, followed by solar radiation metrics. Regional values of these metrics tended to display higher correlations than did local ones, and the counts of extremes temperatures were more strongly correlated than mean temperatures. The only metrics with a strong negative correlation were the precipitation frequency metrics. Neither weather nor air-flow type predictor metrics correlated as well with ozone metrics as the surface climate metrics did. Of the weather metrics, stagnation events showed higher correlations than did the frequency of frontal passages. The 10m level air-flow metrics outperformed their 850 mb counterparts.

Due to weak correlations with ozone, some predictor metrics were regularly excluded from the PCA-selection process. Frontal passages were one such metric. This may reflect a lack of a one-to-one relationship at the *seasonal scale* (i.e. frontal passages during days of low ozone concentrations may not affect the number of days with high ozone concentrations). Soil moisture metrics were another metric group that was not included in the post-correlation phases of the analyses, and this may reflect the weakness of the soil moisture at this particular depth of 40–100 cm as a driver. Atmospheric moisture metrics generally showed weak correlations with the ozone metrics as well. One explanation may be that dewpoint temperature is a poor predictor of ozone at the *seasonal* scale compared to *synoptic* scale.

Overall, the teleconnection predictor metrics showed relatively weak correlations (Fig. 4). The large differences from one predictor metric to the next in the numbers of stations that were selected (via

correlation analysis to participate in the PCA process) for each predictor metric were less apparent than for the meteorological predictors (Table 1). The strongest positive correlations with ozone were observed with the June-July-August (JJA) NAO, JJA PDO, May-June-July-August (MJJA) AO, December-January-February (DJF) QBO and DJF T/NH metrics. The March-April-May (MAM) PDO displayed the only strong negative correlation. Interestingly, although the AMO was positively associated with temperature and weak-negatively correlated with precipitation in the study region, the AMO-based metrics were the most weakly correlated with the ozone metrics.

3.3. Principal Components Analysis and Multiple Linear Regression

PCA and MLR results complemented each other. The former described how often each predictor metric was used in the MLR models, and the latter described the strength of those regression coefficients. The PCA results were in the form of counts of stations whose models employed the predictor metric for *both* ozone metrics. The MLR results were the mean regression coefficient over all stations and both ozone metrics. The spatial distributions of these results were not shown due to little observed spatial structure.

Results of the meteorological predictor metrics (Fig. 5) indicated that the metrics most selected in the PCA were the number of 90th percentile occurrences in regional temperature, mean local daily average solar radiation flux and number of days of regional rain. Overall, the PCA and MLR results were fairly consistent even though the strength of the regression coefficients and frequency with which each metric was selected were occasionally dissimilar.

The results indicated that the most common meteorological metric was temperature based, as anticipated. Interestingly, the strongest metric was the extreme temperature counts and not the mean temperature. This may reflect the relationship that extreme high temperatures have with stagnation and additional biogenic emissions. Plants also remove less ozone from the atmosphere during extremely hot temperatures (Wesley 1989). The strength of the MLR coefficients supported the conclusion of Gégo et al. (2007) that, as a predictor, radiation flux is on par with air temperatures.

The relatively strong correlations between the precipitation metrics and ozone was surprising given that precipitation is not established as a predictor of ozone (Jacob and Winner, 2009). As previously mentioned, the negative correlation with ozone may stem from precipitation acting as a loss mechanism of NO_x through wet deposition of water-soluble reservoir species (e.g. HNO_3). A second possible mechanism may be less direct, in that precipitation tended to be positively correlated with atmospheric moisture yet negatively correlated with solar radiation, daily maximum air temperatures and stagnation. Its selection through the PCA selection process, without exceptionally strong correlations, suggests that the sum of numerous indirect links with ozone may be substantial.

Results differentiating between the regional and local predictor metrics were (to the best of our knowledge) the first to be reported. They indicated that regional metrics displayed modest superiority as predictors, which may reflect the importance of horizontal mixing, climate data resolution or physical processes.

Stagnation-related predictor metrics were rarely selected during the PCA process. This may reflect the paucity of ozone production, and therefore trapping, within the study region. With this logic, urban sites should be more likely to be influenced by stagnation and our results indicated a slight preference towards a correlation between the more urbanized stations and stagnation metrics (not shown).

Air-flow metrics were expected to display closer empirical relationships with ozone, given their previous dependence upon regional transport mechanisms. One explanation was that direction of airflow at a given station was a poor estimate of its origin.

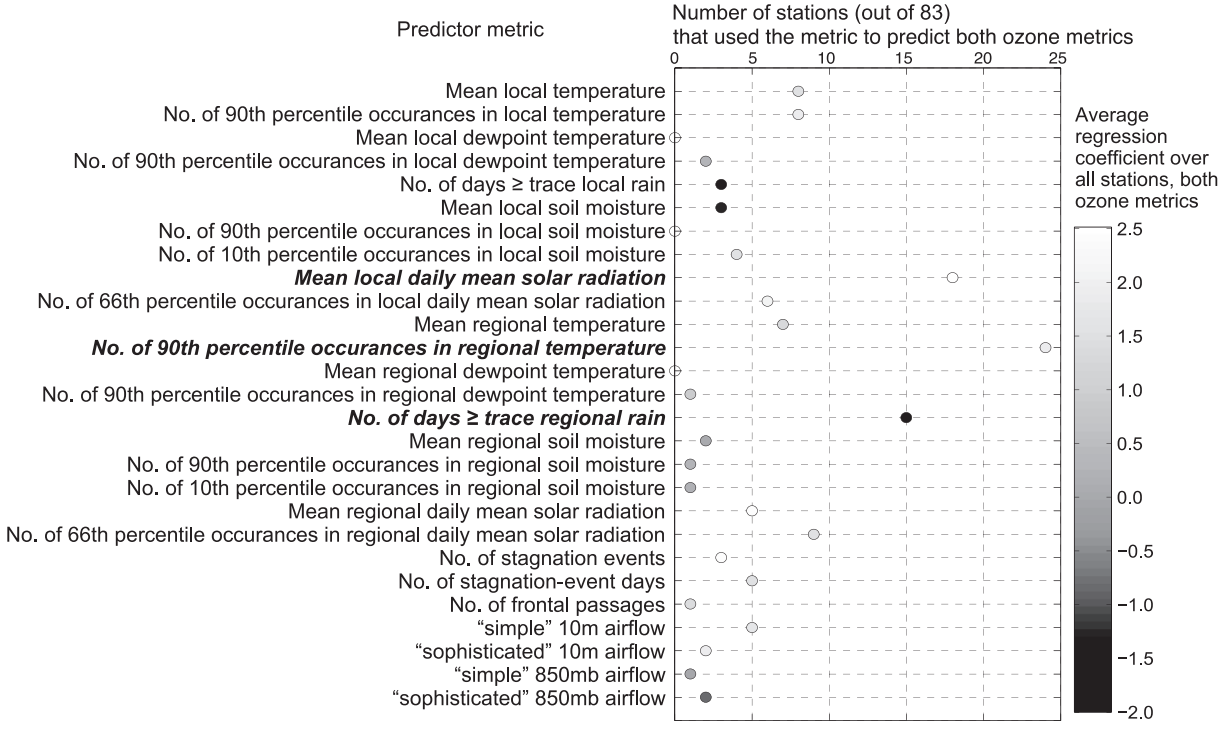

Fig. 5. Number of stations at which each meteorological predictor metric was selected for the linear regression model for both ozone metrics. The average value of the regression coefficient value across all stations and both ozone metrics is shown in choropleth shading.

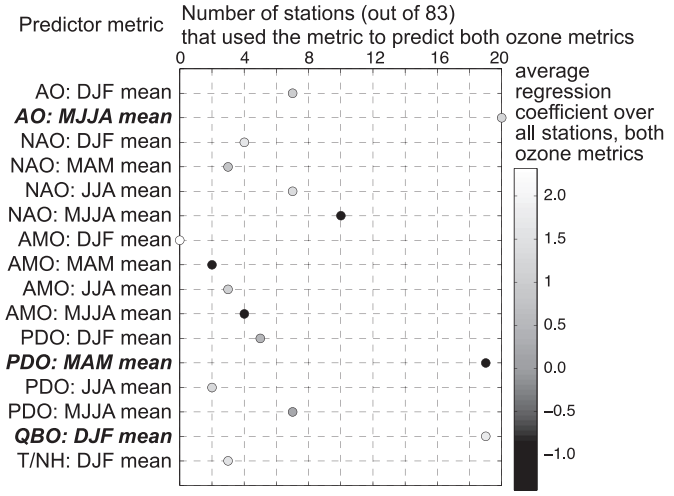

Fig. 6. Number of stations at which each teleconnection predictor metric was selected for the linear regression model for both ozone metrics. The average value of mean regression coefficient value across all stations and both ozone metrics is shown in choropleth shading.

Another factor was that air-flow metrics did not display one-to-one relationships with ozone on the seasonal scale (i.e. air could travel from traditionally polluted regions even when there was no pollution to be advected). Finally, air-flow was often correlated with temperature (a stronger correlated predictor) and thus, may have failed the PCA selection process. In light of these mechanisms, urban sites should be less dependent upon regional transport since ozone could be produced locally; we observed modest preferences for less urbanized stations to correlate with air flow metrics (not shown).

PCA and MLR results of the teleconnection predictor metrics (Fig. 6) showed that while teleconnection-climate indices were overall not strong predictors of ozone, the MJJA AO, MAM PDO and DJF QBO metrics were the strongest examined. In offering viable physical explanations in support of those empirical linkages below, we recommend that they be explicitly investigated in a follow up study.

First, the T/NH predictor metric rarely met the PCA selection criteria despite its statistical associations with DJF temperature in the region. The results suggest the high correlation with the stronger correlated DJF QBO may have weakened its selection in the PCA. Secondly, the MJJA AO could influence the summer ozone concentrations through the strength of the circumpolar vortex. Indeed, the summer AO was positively associated with cool air outbreaks and negatively associated with storminess in the Northeastern U.S. during the summer (Thompson and Wallace, 2000). Thus, through the links with temperature and precipitation, we postulate the MJJA AO's positive correlation with ozone.

We suggest that the MAM PDO's influence on summer ozone concentrations in the Northeastern U.S. could have been through modulation of the Aleutian Low and consequently the position of the Polar Jet Stream position in the Northeastern U.S. The MAM PDO showed a positive statistical association with spring precipitation in the region. We postulate that enhanced springtime precipitation could have led to increased summer soil moisture and therefore lower air temperatures, particularly in the daytime maxima. Since ozone is related to daytime temperatures, it is reasonable to believe that this mechanism may have contributed to the MAM PDO's negative association with summer ozone. McCabe et al. (2004) demonstrated a similar relationship between the PDO and drought over the Northeastern U.S.

Finally, we postulate that the DJF QBO's influence on summer ozone levels in the Northeastern U.S. was via modulation of the DJF polar vortex's strength (Baldwin et al., 2001). The polar vortex controls the probability of cold air outbreaks in the region and thus snow accumulation. The timing of spring snowmelt may then influence the summer temperatures through soil moisture effects

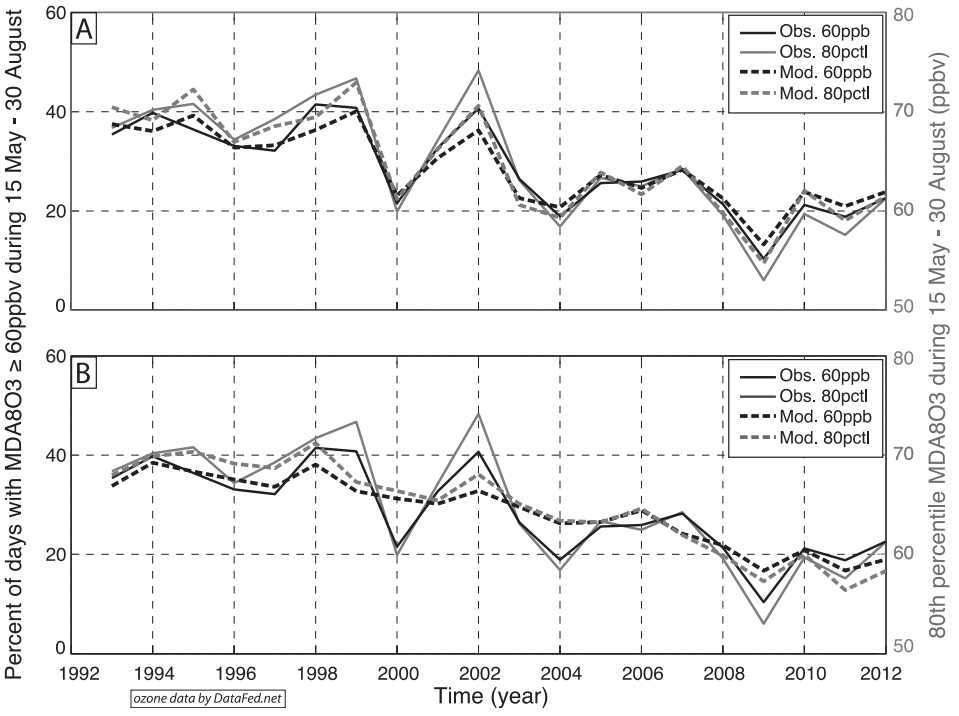

Fig. 7. Time series of the observations and regression model output for the percent of days with MDA8O3 ≥60 ppbv 15 May – 30 August over the 1993–2012 period for the 83 study stations. Panel A shows the meteorological-based predictor metric model, and Panel B the teleconnection-based predictor metric model.

Table 3
Performance of the regression models[a].

	Cross validation		Post building	
	60 ppbv	80 pctl	60 ppbv	80 pctl
Meteorological-based metrics model				
R^2	0.57 (0.17)	0.62 (0.17)	0.74 (0.10)	0.76 (0.10)
RMSE ratio	0.64 (0.13)	0.60 (0.13)	0.50 (0.10)	0.48 (0.10)
MAD ratio	0.64 (0.20)	0.59 (0.21)	0.50 (0.17)	0.47 (0.17)
Teleconnection-based metrics model				
R^2	0.39 (0.21)	0.38 (0.19)	0.59 (0.13)	0.58 (0.13)
RMSE ratio	0.77 (0.13)	0.78 (0.12)	0.63 (0.11)	0.64 (0.10)
MAD ratio	0.73 (0.25)	0.73 (0.24)	0.60 (0.21)	0.59 (0.19)

[a] Based on ability of regression models, in conjunction with NOx *retrending*, in recreating the time series both ozone metrics during 15 May – 30 August from 1993 to 2012; statistical metrics were averaged over the 83 study stations; standard deviations are provided in parentheses.

(Westerling et al., 2006). Since summertime temperatures were positively associated with ozone, the net result would match a positive correlation between the DJF QBO and summer ozone concentrations.

3.4. Quality of the models

The 83-station mean time series of ozone was recreated via output of the individual models (Fig. 7). Meteorological-based models were clearly better at replicating the ozone metrics than the teleconnection-based models. The teleconnection-based models had a notably lower interannual variability than was observed.

Results of model performance demonstrated agreement between the ozone metrics and RMSE and MAD ratios (Table 3). The meteorological-based models performed well, with roughly between a 0.60–0.75 R^2 and RMSE/MAD ratios of 0.50–0.65. The teleconnection-based models were inferior, with R^2 values roughly between 0.40 and 0.60 and RMSE/MAD ratios between 0.60 and 0.75 (see Table 3).

4. Conclusions

Meteorological-based MLR models performed well enough to become practical tools for predicting ozone metrics, as long as accurate forecasts can be acquired. The strongest meteorological predictors were the frequency of regional extreme temperatures, local solar radiation flux, and regional precipitation frequency. There was a tendency for regional variables to be better predictors than the analogous local variables. Dewpoint temperatures, frontal passages, air-flow patterns and soil moisture had substantially little predictive power. Although teleconnection-based models did not accurately predict the ozone metrics, they were successful enough to suggest linkages with ozone and therefore warrant further investigation. These links with the global climate system were through the MJJA AO, the MAM PDO and the DJF QBO. Our results suggest physical pathways by which these teleconnections may influence ozone through temperature and precipitation modulation.

Implications of this work predominantly pertain to climate controls on ozone in the Northeastern U.S. Our results suggest that when anticipating upcoming ozone-seasons, agencies concerned with ozone exceedences should shift their attention to the climate outlooks of extreme temperatures, cloudiness and precipitation frequency. Future studies might be more inclined to include precipitation metrics; use regional meteorological information instead of local; and use the frequency of temperature extremes instead of mean temperatures. Other implications of our work pertain to the methodology for extracting large numbers of predictors to small sets of quasi-independent predictors. We have found that variable filtering via correlation analysis followed by PCA with Varimax Rotation to be a suitable approach.

The next steps in this research would focus on the impacts of climate change on the frequency of high ozone levels using future climate projections with these meteorological-based models. Longer time periods would be beneficial in confirming these relationships with Northeastern U.S. ozone. The relationship between particulate matter and meteorological/climate information should also be explored. Finally, it would be informative to confirm the roles played by teleconnections in influencing ozone in the Northeastern U.S. by a robust examination (e.g. model simulations) of the physical linkages involved.

Acknowledgments

Funding for this work was provided by the National Science Foundation (NSF GEO-1341620), Vermont Agency of Natural Resources, the Vermont Department of Health and the New York State/Union of University Professions Individual Development Award Program (IDAP) through SUNY Plattsburgh.

References

Baldridge, E., 2005. Evaluating Ozone Control Programs in the Eastern United States: Focus on the NOx Budget Trading Program, 2004. United States Environmental Protection Agency.

Baldwin, M.P., Gray, L.J., Dunkerton, T.J., Hamilton, K., Haynes, P.H., Randel, W.J., Holton, J.R., Alexander, M.J., Hirota, I., Horinouchi, T., Jones, D.B.A., Kinnersley, J.S., Marquardt, C., Sato, K., Takahashi, M., 2001. The Quasi-Biennial Oscillation. Rev. Geophys. 39, 179–229. http://dx.doi.org/10.1029/1999RG000073.

Bloomer, B.J., Stehr, J.W., Piety, C.A., Salawitch, R.J., Dickerson, R.R., 2009. Observed relationships of ozone air pollution with temperature and emissions. Geophys. Res. Lett. 36, L09803. http://dx.doi.org/10.1029/2009GL037308.

Chan, E., 2009. Regional surface-level ozone trends in the context of meteorological influences across Canada and the Eastern United States from 1997–2006. J. Geophys. Res. 114, D05301. http://dx.doi.org/10.1029/2008JD010090.

Crutzen, P.J., Lawrence, M.G., 2000. The impact of precipitation scavenging on the transport of trace gases: a 3-dimensional model sensitivity study. J. Atmos. Chem. 37, 81–112. http://dx.doi.org/10.1023/A:1006322926426.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Paseris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064. http://dx.doi.org/10.1002/joc.1688.

Enfield, D.B., Mestas-Nunez, A.M., Trimble, P.J., 2001. The Atlantic multidecadal oscillation and it's relation to rainfall and river flows in the continental U.S. Geophys. Res. Lett. 28, 2077–2080. http://dx.doi.org/10.1029/2000GL012745.

Fischer, E.M., Seneviratne, S.I., 2007. Soil moisture-atmosphere interactions during the 2003 European summer heat wave. J. Clim. 20, 5081–5099. http://dx.doi.org/10.1175/JCLI4288.1.

Frey, H. C. Chair, Clean Air Scientific Advisory Committee (CASAC) Letter to G. McCarthy, EPA Administrator, CASAC Review of the EPA's Second Draft Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards, EPA-CASAC-14–004, June 26, 2014. http://yosemite.epa.gov/sab/sabproduct.nsf/5EFA320CCAD326E885257D030071531C/$File/EPA-CASAC-14-004+unsigned.pdf.

Gégo, E., Porter, P.S., Gilliland, A., Rao, S.T., 2007. Observation-based assessment of the impact of nitrogen oxides emissions reductions on ozone air quality over the eastern United States. J. Appl. Meteorol. Climatol. 46, 994–1008. http://dx.doi.org/10.1175/JAM2523.1.

Husar, R.B., Rendard, W.P., 1997. Ozone as a Function of Local Wind Direction and Wind Speed: Evidence of Local and Regional Transport. Center for Air Pollution Impact and Trend Analysis (CAPITA) report. May 1997.

Husar, R.B., Hoijarvi, K., Falke, S.R., Robinson, E.M., Percivall, G.S., 2008. DataFed: an architecture for federating atmospheric data for GEOSS. Syst. J. IEEE 2, 366–373. http://dx.doi.org/10.1109/JSYST.2008.2003292.

Jacob, J.D., Winner, D.A., 2009. Effect of climate change on air quality. Atmos. Environ. 43, 51–63. http://dx.doi.org/10.1016/j.atmosenv.2008.09.051.

Lamarque, J.-F., et al., 2010. Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. Atmos. Chem. Phys. 10, 7017–7039. http://dx.doi.org/10.5194/acp-10-7017-2010.

LeClair, Vincent, 1997. OTAG recommends ozone controls tailored to pollution transport. Environ. Sci. Technol. 31, 352A–353A.

Leibensperger, E.M., Mickely, L.J., Jacob, D.J., 2008. Sensitivity of U.S. air quality to

mid-latitude cyclone frequency and implications of 1980–2006 climate change. Atmos. Chem. Phys. 8, 7075–7086. http://dx.doi.org/10.5194/acp-8-7075-2008.
Lin, M., Fiore, A.M., Horowitz, L.W., Langford, A.O., Oltmans, S.J., Tarasick, D., Harald, H.E., 2015. Climate variability modulates western U.S. ozone air quality in spring via deep stratospheric intrusions. Nat. Commun.
Lin, M., Horowitz, L.W., Oltmans, S.J., Fiore, A.M., Songmiao, F., 2014. Tropospheric ozone trends at Manna Loa Observatory tied to decadal climate variability. Nat. Geosci. 7, 136–143. http://dx.doi.org/10.1038/NGEO2066.
Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., Francis, R.C., 1997. A pacific interdecadal climate oscillation with impacts on salmon production. Bull. Am. Meteorol. Soc. 78, 1069–1079. http://dx.doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.
McCabe, G.J., Palecki, M.A., Betancourt, J.L., 2004. Pacific and Atlantic ocean influences on multidecadal drought frequency in the United States. PNAS 101, 4136–4141. http://dx.doi.org/10.1073/pnas.0306728101.
Meinshausen, M., et al., 2011. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Clim. Change 109, 213–241. http://dx.doi.org/10.1007/s10584-011-0156-z.
Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. 87, 342–360. http://dx.doi.org/10.1175/BAMS-87-3-343.
Milanchus, M.L., Rao, S.T., Zurbenko, I.G., 1998. Evaluating the effectiveness of ozone management efforts in the presence of meteorological variability. J. Air Waste Manag. 48, 201–215. http://dx.doi.org/10.1080/10473289.1998.10463673.
Mo, K.C., Livezey, R.E., 1986. Tropical–extratropical geopotential height teleconnections during the northern hemisphere winter. Mon. Weather Rev. 114, 2488–2515. http://dx.doi.org/10.1175/1520-0493(1986)114<2488:TEGHTD>2.0.CO;2.
Rao, S.T., et al., 2003. Summertime characteristics of the atmospheric boundary layer and relationships to ozone levels over the Eastern United States. Pure Appl. Geophys. 160, 21–55. http://dx.doi.org/10.1007/s00024-003-8764-9.
Rasmussen, D.J., Fiore, A.M., Naik, V., Horowitz, L.W., McGinnis, S.J., Schultz, M.G., 2012. Surface ozone-temperature relationship in the Eastern US: a monthly climatology for evaluating chemistry-climate models. Atmos. Environ. 47, 142–153. http://dx.doi.org/10.1016/j.atmosenv.2011.11.021.
Schichtel, B.A., Husar, R.B., 2001. Eastern North American transport climatology during high- and low- ozone days. Atmos. Environ. 35, 1029–1038. http://dx.doi.org/10.1016/S1352-2310(00)00370-8.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of the CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498. http://dx.doi.org/10.1175/BAMS-D-11-00084.1.
Thompson, D.W.J., Wallace, J.M., 2000. Annular modes in the extratropical circulation. Part 1: month-to-month variability. J. Clim. 13, 1000–1016. http://dx.doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.
Thompson, M.L., Reynolds, J., Cox, L.H., Guttorp, P., Sampson, P.D., 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. Atmos. Environ. 35, 617–630. http://dx.doi.org/10.1016/S1352-2310(00)00261-2.
Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190, 214–251. http://dx.doi.org/10.1016/S0022-1694(96)03128-9.
Valente, R.J., Imhoff, R.E., Tanner, R.L., Meagher, J.F., Daum, P.H., Hardesty, R.M., Banta, R.M., Alvarez, R.J., McNider, R.T., Gillani, N.V., 2012. Ozone production during an urban air stagnation episode over Nashville, Tennessee. J. Geophys. Res. Atmos. 103, 22555–22568. http://dx.doi.org/10.1029/98JD01641.
Vukovich, F.M., 1994. Boundary layer ozone variations in the Eastern United States and their associations with meteorological variations: long-term variations. J. Geophys. Res. 99, 16839–16850. http://dx.doi.org/10.1029/93JD02554.
Vukovich, F.M., 1995. Regional-scale boundary layer ozone variations in the Eastern United States and their association with meteorological variables. Atmos. Environ. 29, 2259–2273.
Wang, J.X.L., Angell, J.K., 1999. Air Stagnation Climatology for the United States (1948–1998). NOAA/Air Resources Laboratory ATLAS No. 1.
Wesely, M.L., 1989. Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical-models. Atmos. Environ. 23, 1293–1304. http://dx.doi.org/10.1016/0004-6981(89)90153-4.
Westerling, A.L., Hidalgo, H.G., Cayan, D.R., Swetnam, T.W., 2006. Warming and earlier spring increase Western U.S. forest wildfire activity. Science 313, 941–943. http://dx.doi.org/10.1126/science.1128834.