Journal of the Air & Waste Management Association

ISSN: 1096-2247 (Print) 2162-2906 (Online) Journal homepage: http://www.tandfonline.com/loi/uawm20

Comparing Neural Networks and Regression Models for Ozone Forecasting

Andrew C. Comrie

To cite this article: Andrew C. Comrie (1997) Comparing Neural Networks and Regression Models for Ozone Forecasting, Journal of the Air & Waste Management Association, 47:6, 653-663, DOI: 10.1080/10473289.1997.10463925

To link to this article: https://doi.org/10.1080/10473289.1997.10463925

TECHNICAL PAPER

ISSN 1047-3289 *J. Air & Waste Manage. Assoc.* **47**: 653-663

Comparing Neural Networks and Regression Models for Ozone Forecasting

Andrew C. Comrie
University of Arizona, Tucson, Arizona

ABSTRACT

Many large metropolitan areas experience elevated concentrations of ground-level ozone pollution during the summertime "smog season". Local environmental or health agencies often need to make daily air pollution forecasts for public advisories and for input into decisions regarding abatement measures and air quality management. Such forecasts are usually based on statistical relationships between weather conditions and ambient air pollution concentrations. Multivariate linear regression models have been widely used for this purpose, and well-specified regressions can provide reasonable results. However, pollution-weather relationships are typically complex and nonlinear¾especially for ozone¾properties that might be better captured by neural networks. This study investigates the potential for using neural networks to forecast ozone pollution, as compared to traditional regression models. Multiple regression models and neural networks are examined for a range of cities under different climate and ozone regimes, enabling a comparative study of the two approaches. Model comparison statistics indicate that neural network techniques are somewhat (but not dramatically) better than regression models for daily ozone prediction, and that all types of models are sensitive to different weather-ozone regimes and the role of persistence in aiding predictions.

IMPLICATIONS

Daily ozone forecasts by metropolitan air-quality agencies are commonly made using multiple regression models, which are based on linear relationships between weather data and other pollution-related variables. Neural network models are not yet widely used for this purpose, but they have the potential to incorporate complex, nonlinear relationships such as those controlling ozone formation. A series of direct comparisons for eight cities shows that neural networks can provide modest improvements over regression-based ozone predictions, and that the performance of either type of model is improved by including persistence information.

INTRODUCTION

Weather-Ozone Relationships

The photochemical processes leading to ozone formation are complex and nonlinear, and they are well-documented.[1-3] Ozone (O_3) is a secondary pollutant, in that it is not usually emitted directly from tailpipes or smokestacks, but instead is formed in the atmosphere as a result of reactions between other directly emitted pollutants. These primary pollutants (ozone precursors) result from the use of gasoline, other petrochemicals, and fossil fuels, and they are emitted largely by industry and automobiles. Ozone precursors fall into two related groups: various oxides of nitrogen (NO_x), and a range of volatile organic compounds (VOCs) such as evaporative solvents and other hydrocarbons. In suitable ambient meteorological conditions (e.g., a warm, sunny/clear day), ultraviolet radiation (UV) causes the precursors to interact photochemically in a set of reactions that result in the formation of ozone (and several other photochemical pollutants). Ozone is simultaneously destroyed via oxidation of an additional reactant (D), typically one of the NO_x. These relationships can be expressed conceptually as

$$NO_x + VOCs + UV \rightarrow O_3 \quad (1a)$$

$$O_3 + D \rightarrow DO + O_2 \quad (1b)$$

where the roles of catalysts and minor chemical species are ignored. Each term in these reactions relates to the weather-dependency of ozone production in several respects:

- Precursor emissions processes (combustion, evaporation) are sensitive to temperature, vapor pressure, etc.
- Dispersion conditions (e.g., wind speed, atmospheric stability, mixing height) influence ambient concentrations of precursors and ozone.
- Reaction rates between the pollutants vary according to temperature and meteorological controls on UV availability (e.g., cloudiness and other forms of attenuation in the atmosphere).

In addition, many of the reactions between pollutants are highly nonlinear; notable examples are the dependence of ozone concentration on the ratio between certain NO_x (photostationary state), and the simultaneous dependence of ozone on VOC concentration.[1-3] Many of the meteorology-pollution interactions are nonlinear, too. Figure 1 illustrates an elementary example of nonlinear weather-related interactions, the combined relationship between total hours of sunshine and daily maxima of temperature and ozone for Boston, MA. For this simple empirical example (in which ozone is likely influenced by more than just the two selected variables), a polynomial (quadratic) regression provides a somewhat better fit to the curved scatter of data points than a linear regression. Not only are there numerous direct links between weather conditions and the ozone production factors mentioned above, but there are also equally important feedbacks among the many variables. The net result is a highly complex system of ozone formation mechanisms that displays the compounded effects of multiple chemical and meteorologically related nonlinearities.

Ozone Forecasting

The intricacies of ozone formation make day-to-day operational prediction of ozone quite difficult. One of the best ways to capture these complex interactions is through the use of photochemical models used mainly for research and planning (e.g., the Urban Airshed Model). But such models are unsuitable in many operational settings because they require significant computer and staffing commitments, as well as many complex chemical and meteorological inputs (precursor concentrations, mesoscale meteorological measurements, etc.). Thus, while such models are theoretically sophisticated and desirable for forecasting, they are not practical choices in many locations. The most common alternative is to employ a multivariate statistical approach, which is widely used in operational ozone forecasting and research-oriented statistical modeling.[4-11] Multiple (multivariate) linear regression is the most popular of these techniques, and it has the general form:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_i \qquad (2)$$

where, for a set of i successive observations, the predictand variable Y is a linear combination of an offset β_0, a set of k predictor variables X with matching β coefficients, and a residual error ε. The β values are commonly derived via the procedure of ordinary least squares (although there are other methods). When the regression equation is used in predictive mode, ε (the difference between actual and predicted values not accounted for by the model) is omitted because its expected value is zero. Note that regression models are inherently linear, although curvilinear relationships can be incorporated via polynomial terms in the regression, and known relationships can be pre-specified by transforming a nonlinear predictor variable into a more linear form (e.g., by taking a logarithm) before using it in the model.

Regression models of ozone pollution typically incorporate from one or two input variables[7] to as many as 313 variables (reflecting a range of weather data from several atmospheric levels that are potentially correlated with ozone concentrations).[8] A stepwise multiple regression procedure is commonly used to produce a parsimonious model that maximizes accuracy with an optimally reduced number of predictor variables. In most locations, temperature is the meteorological variable most highly correlated with ozone, although wind speed, variables related to UV, and atmospheric moisture are sometimes included in regression models.

Because daily maximum ozone concentrations are partially dependent on the previous day's concentrations, ozone data display strong serial correlation. In contrast, regression models assume that observations are statistically independent events. To avoid this problem while still incorporating the importance of persistence, some investigators have used a lagged ozone concentration (typically a value from the previous day) as an additional predictor variable in the model.[7,8] If the regression is performed in an explanatory mode (for interpretation of coefficients or significance testing), then this strategy is statistically somewhat awkward, because the inclusion of serially correlated ozone data does not necessarily aid understanding of weather-ozone relationships. In contrast,

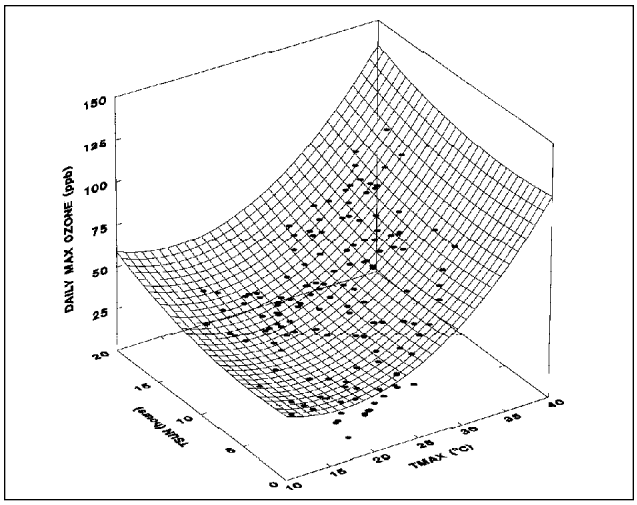

Figure 1. An empirical example of the nonlinear relationships between total hours of sunshine (TSUN) and daily maxima of temperature (TMAX) and ozone for the Boston, MA, site, May-September 1995. The best-fit quadratic regression surface ($R^2 = 0.65$) is represented by MAX. OZONE = 72.73 - 5.05 TMAX - 0.57 TSUN + 0.16 $TMAX^2$ + 0.09 $TSUN^2$ - 0.01 (TMAX x TSUN), while the best-fit linear regression plane $R^2 = 0.55$ (not illustrated) is represented by MAX. OZONE = -10.77 + 2.30 TMAX + 0.50 TSUN.

the inclusion of lagged data in regression modeling is desirable when used in a predictive mode, frequently improving the accuracy of predictions.

Neural Networks

Neural networks (or, properly, artificial neural networks) are loosely based on biological neural systems, in that they are made up of an interconnected system of nodes (neurons). Also, a neural network can identify patterns in numeric data in a somewhat analogous fashion to the learning process in its biological counterpart (i.e., by accumulated experience, or training). Neural networks are highly robust with respect to underlying data distributions (non-parametric), and no assumptions are made about relationships between variables (unlike the linear or pre-specified curvilinear relationships in regression). Thus, neural networks are well-suited to modeling complex, nonlinear phenomena[12] such as the stock market[13] or tornado formation.[14] To date, neural networks have received limited application to the problem of air pollution forecasting,[15,16] and no studies comparing them to regression techniques have been performed.

There are many neural network models, but the basic structure involves a system of layered, interconnected nodes (Figure 2). The nodes are arranged to form an input layer, one or more "hidden" layers, and an output layer, with nodes in each layer connected to all nodes in neighboring layers. Nodes in the hidden and output layers comprise several elements: a set of numerical inputs that convey information to the node; a matching set of weights (one for each input, plus an optional bias weight); activation and transfer functions that transform the weighted inputs; and a numerical output. The input layer supplies data to the hidden layer and does not contain activation or transfer functions. A typical feed-forward network (as in Figure 2, with notation as illustrated) might use a dot-product activation function that, for each node B_j (j = 1, 2, … , n) in the hidden layer, is computed as:

$$B_j = \sum_{i=1}^{m} w_{ij} A_i + w_{0j} A_0 \qquad (3)$$

with input nodes A_i (i = 1, 2, … , m) and weights w_{ij} between nodes A_i and B_j. The bias node (A_0) typically has a constant input of 1, with a matching weight w_{0j}. A similar calculation is made for each node C_k (k = 1, 2, … , o) in the output layer (o = 1 for the example in Figure 2), using weights w_{jk} between nodes B_j and C_k (with w_{0k} and B_0 for the bias). Each node value is subsequently passed through a transfer function, which may

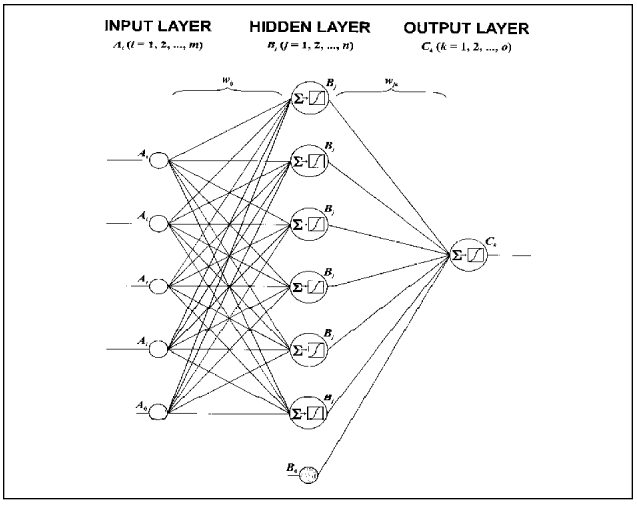

Figure 2. Schematic example of an m x n x o artificial neural network, showing a multilayer perceptron with a 4 x 6 x 1 structure (additional shaded circles indicate bias nodes). Information enters the network at the input layer nodes (A_i) and moves along weighted links (w_{ij} and w_{jk}) to nodes in the hidden and output layers (B_j and C_k), which each contain an activation function (∑) and a nonlinear transfer function (see text for details). Each node combines information from all nodes in the previous layer, resulting in a final output that is a complex function of inputs and internal network transformations.

be linear or nonlinear. A common choice of nonlinear transfer function is a sigmoid, of the general form:

$$\sigma(u) = \left[1 + e^{-u}\right]^{-1} \qquad (4)$$

where $u = B_j$ (or C_k). Nonlinearities are incorporated into the network via the activation and transfer functions in each node. Complexities in the data are captured through the number of nodes in the hidden layers. The weights are determined by iteration to produce the lowest error in the output, measured, for example, as the root mean squared error (RMSE):

$$\mathrm{RMSE} = \left\{\left[\sum_{l=1}^{p}\sum_{k=1}^{o}\left(C_{kl} - T_{kl}\right)^2\right] p^{-1}\right\}^{0.5} \qquad (5)$$

where C_{kl} is the (predicted) neural network output for training pattern l (i.e., input observations A_i) of p training patterns, and T_{kl} is the matching (observed) training pattern output. Initial weights are randomly assigned; in subsequent iterations, individual weights are incrementally adjusted to reduce error (i.e., a backpropagation algorithm).[17] To avoid overfitting to the data, a neural network is usually trained on a subset of inputs and outputs to determine weights, and subsequently validated on the remaining (quasi-independent) data to measure the accuracy of predictions. While neural networks can be used

Table 1. Ozone monitors for the eight study sites.

City	AIRS ID #	Site Address
Atlanta, GA	13-089-0002	DeKalb Jr. College
Boston, MA	25-025-1003	Powder Horn Hill
Charlotte, NC	37-119-0034	Plaza and Lakedell
Chicago, IL	17-031-0037	5358 N. Ashland
Phoenix, AZ	04-013-3002	1845 E. Roosevelt
Pittsburgh, PA	42-003-0008	BAPC, 301 39th St.
Seattle, WA	53-033-0010	Lake Sammamish S.P.
Tucson, AZ	04-019-0002	151 W. Congress

in regression-like applications, they are also employed in clustering and classification applications. It is beyond the scope of this paper to review the extensive body of literature on neural networks, but further theoretical and technical details may be found in one of many introductory texts.[12,17,18]

Key Questions

Ozone formation is a complex, nonlinear process that neural networks might be able to capture without many of the usual limiting assumptions of other statistical methods. The central goal of this study, therefore, is to evaluate whether neural networks might be an improvement over conventional regression approaches to ozone forecasting for polluted cities. There are several research questions associated with this goal. First, how do the two modeling approaches perform across different climate and ozone regimes? Even within the United States, the meteorological factors controlling ozone vary greatly from one city to another, thereby affecting the nature and accuracy of the models. Second, how useful or necessary are lagged ozone data as input to the models? Neural networks might perform well without having to employ ozone concentrations from the previous day, or they might at least be less reliant on the use of lagged ozone as a predictor. Third, how well are high-ozone episodes predicted? Regression models generally underpredict the upper tail of the distribution, yet these ozone episodes are often the most important values to forecast when a model is used operationally. Stated less formally, the broad aim of this paper is to see if neural networks are "better" than multiple regression for ozone forecasting, and to see if any improvements over the simpler and more widely understood regression methods are worth the extra effort.

DATA AND METHODS

Ozone Data

Data from eight cities around the United States were used to compare regression models and neural networks under a variety of climate and ozone regimes (see Table 1 for a list of cities). Selection of individual cities, although somewhat arbitrary, was based on optimizing data availability for ozone and weather, obtaining a range of city sizes (as a surrogate for pollutant loading and behavior), and covering a range of climatic environments (some similar, some different). The ozone data used were the daily maximum 1-hour concentrations, for the months of May-September, over the five-year period 1991-1995. The data were obtained from the U.S. Environmental Protection Agency's (EPA) Aerometric Information and Retrieval System (AIRS) database. A single ozone monitoring site was chosen for each city, based primarily on the completeness of the ozone record for the study period, and its location in or near an urban or suburban area (rural sites were avoided because of long-range transport complications). Details of the eight study sites are listed in Table 1. Note that the highest ozone concentrations in any one city are not necessarily recorded at the respective site, but the site is likely to be representative of day-to-day (meteorologically related) fluctuations in ozone that the models are intended to capture. The top panel in Figure 3 illustrates an example of such fluctuations, with a time series of daily 1-hour ozone maxima in Boston for the summer of 1995.

Weather Data

The meteorological data used in the study were extracted from the First-Order Summary of Day (FSOD) database, compiled from standard airport observations by the National Climatic Data Center. The FSOD variables selected for this study were those most closely related to ozone behavior: daily maximum temperature (TMAX), average daily dewpoint temperature (DPTP), average daily wind speed (AWND), and daily total sunshine (TSUN). Maximum temperatures correlate highly with ozone in many locations, and high dewpoint temperatures are strongly associated with stagnating anticyclones responsible for high ozone events, as are low wind speeds and total sunshine (a proxy for UV flux).[19,20,21] The weather data were extracted for the same period as the ozone data (May-September of 1991-1995). Based on the results from other studies,[7,8] it was decided that this study would use a conservative number (four) of meteorological variables in order to maintain parsimony and keep the resulting models simple enough for meaningful comparison. The lower four panels in Figure 3 contain time series of weather data for the summer of 1995 in Boston, matching the ozone data in the top panel. Varying degrees of correspondence between the weather and ozone data can be seen, such as the typical high-ozone conditions found on the episode day

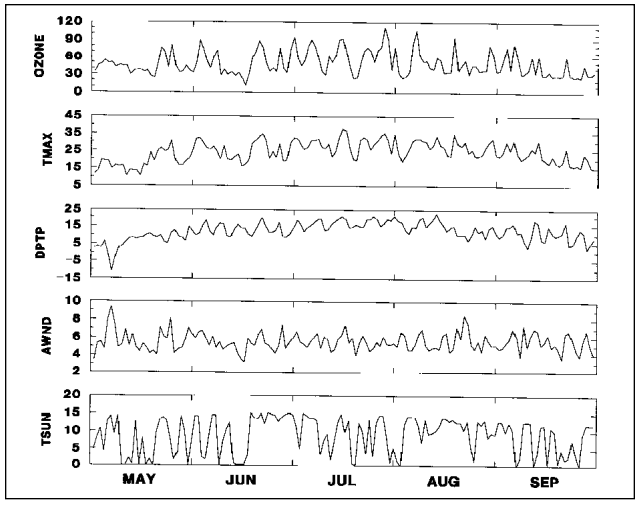

Figure 3. Example time series of daily maximum 1- hour ozone (ppb) and meteorological variables for Boston, MA, 1995 (TMAX and DPTP in °C, AWND in m s^{-1}, and TSUN in hours).

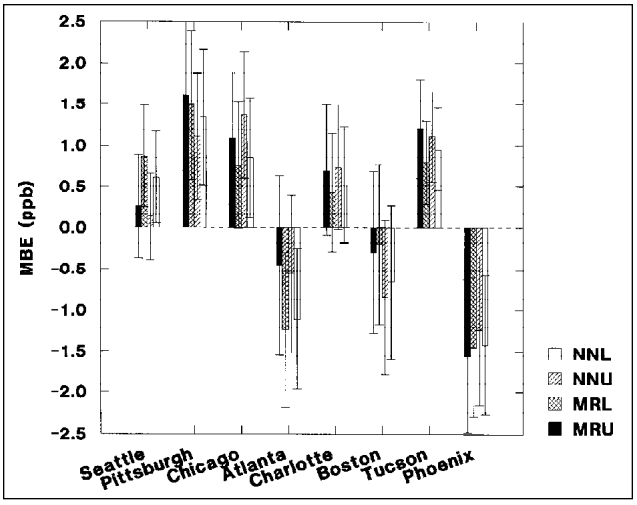

Figure 4. Mean bias errors (average residuals) and associated bootstrap estimates of standard error, showing over- or under-prediction by model and city.

in late July, which displays coincident high temperature, high dewpoint temperature, moderate wind speed and high sunshine. Note the complexity of relationships between the weather variables and ozone, which the models aim to capture, as well as inconsistencies (perhaps due to precursor emissions variability, seasonal effects, or a stochastic component) that the models attempt to overcome. For a scatterplot comparison of some of these time series data (ozone, TMAX, TSUN), refer to Figure 1.

Database Construction

To facilitate comparison of models between cities, a "uniform" database was created from the ozone and weather data described above. Lagged ozone data (the 1-hour ozone maximum from the previous day) were included as an optional fifth predictor variable for each city. Days with errors or that were missing data for any variable were removed from the database. The resulting data completeness for the eight cities ranged between 90 and 98%. All data were scaled to provide values between 0.2 and 0.8 as follows:

$$x_i = 0.2 + 0.6(O_i - O_{min})(O_{max} - O_{min})^{-1} \quad (6)$$

where individual scaled observations (x_i) are produced by standardizing actual observations (O_i) by the minimum (O_{min}) and maximum (O_{max}) ozone concentrations for each city. Scaling was performed for two reasons: to provide commensurate data ranges so that the regressions were not dominated by any variable that happened to be expressed in large numbers; and, to avoid the (essentially flat) asymptotes of the sigmoid function (a common practical step in neural network modeling). After modeling, but prior to model comparisons, scaled ozone predictions produced by the models (y_i) were reverse-scaled using:

$$P_i = O_{min} + 0.6^{-1}(y_i - 0.2)(O_{max} - O_{min}) \quad (7)$$

which returns individual predictions in the original units (P_i in ppb).

Test Methodology

Prior to model development and testing, the chronological order of each city's data was randomized. A total of 690 observations were used for each of the eight cities. Development of regression models and training of the neural networks were carried out on a subset of 440 (random) observations for each city; the remaining (random) 250 observations were reserved as a quasi-independent subset for regression validation and network testing.

The overall rationale for the model comparison was to keep things as simple as possible. Regression models and neural networks are quite different in nature, and both can be adjusted or made more sophisticated to suit an individual application. For ease of comparison it was decided to avoid such complications, and to implement the models in their most straightforward, typical fashion. Hypothetically, then, contrasts in the amount of useful information extracted from the same basic set of weather-ozone relationships (i.e., model performance) would be due to intrinsic differences between the basic models.

Following this rationale, "standard" multiple regression modeling was performed, without transforming any input variables or employing a stepwise variable-selection procedure. The latter was done to maximize explained variance (R^2) and to keep the results in comparable form, thereby offsetting the minor loss in

Table 2. Model comparison statistics for the 250 quasi-independent test observations (MRU = multiple regression unlagged; NNU = neural network unlagged; MRL = multiple regression lagged; NNL = neural network lagged). Units for b, d_1, d_2, and R^2 are dimensionless; all other measures are in ppb. Bootstrap estimates of standard error are listed in parentheses (those for $RMSE_s$ and $RMSE_u$ are not independent, and are included for comparison only).

Statistic	Model	Seattle	Pittsburgh	Chicago	Atlanta	Charlotte	Boston	Tucson	Phoenix
$\bar{O}$	-	41.07 (0.99)	56.75 (1.41)	39.50 (1.15)	66.49 (1.42)	64.44 (1.22)	49.62 (1.24)	53.31 (0.63)	62.89 (1.02)
$\bar{P}$	MRU	41.33 (0.82)	58.37 (1.10)	40.59 (0.82)	66.04 (1.13)	65.14 (0.83)	49.33 (0.78)	54.51 (0.28)	61.34 (0.56)
$\bar{P}$	MRL	41.94 (0.80)	58.24 (1.20)	40.25 (0.86)	65.26 (1.11)	64.87 (0.96)	49.42 (0.84)	54.11 (0.42)	61.44 (0.70)
$\bar{P}$	NNU	41.20 (0.84)	57.86 (1.17)	40.87 (0.89)	65.93 (1.16)	65.18 (0.91)	48.78 (0.79)	54.42 (0.34)	61.65 (0.51)
$\bar{P}$	NNL	41.68 (0.81)	58.09 (1.21)	40.35 (0.90)	65.38 (1.14)	64.96 (1.00)	48.96 (0.85)	54.26 (0.41)	61.47 (0.67)
s_o	-	16.51 (0.93)	24.07 (1.11)	18.93 (1.16)	25.54 (1.38)	20.85 (0.78)	21.03 (0.93)	11.42 (0.46)	16.91 (0.80)
s_p	MRU	13.80 (0.57)	18.33 (0.77)	13.77 (0.55)	16.85 (0.97)	14.11 (0.68)	13.30 (0.54)	5.28 (0.31)	9.55 (0.53)
s_p	MRL	13.68 (0.59)	19.10 (0.76)	14.53 (0.60)	18.29 (0.98)	15.47 (0.62)	14.41 (0.54)	7.80 (0.46)	11.11 (0.57)
s_p	NNU	14.36 (0.70)	19.14 (0.82)	15.12 (0.62)	18.98 (0.87)	15.27 (0.62)	13.60 (0.48)	6.30 (0.34)	8.79 (0.35)
s_p	NNL	13.86 (0.70)	19.17 (0.73)	15.28 (0.63)	19.67 (0.95)	16.14 (0.60)	14.49 (0.51)	7.69 (0.34)	10.63 (0.48)
a	MRU	15.03 (1.60)	25.68 (2.30)	19.53 (1.53)	33.84 (2.92)	33.55 (2.35)	29.97 (1.45)	44.97 (1.70)	44.22 (2.09)
a	MRL	14.76 (1.52)	22.44 (2.30)	16.73 (1.53)	28.61 (2.53)	28.24 (2.41)	26.47 (1.52)	30.73 (2.05)	36.29 (2.37)
a	NNU	11.29 (1.30)	21.61 (1.82)	16.27 (1.26)	27.92 (2.47)	29.98 (2.34)	26.74 (1.33)	39.14 (1.88)	45.69 (1.72)
a	NNL	13.03 (1.36)	21.14 (1.93)	14.82 (1.38)	24.62 (2.20)	26.12 (2.43)	24.37 (1.35)	30.54 (1.78)	37.49 (2.13)
b	MRU	0.64 (0.04)	0.58 (0.03)	0.53 (0.04)	0.49 (0.04)	0.49 (0.03)	0.39 (0.02)	0.18 (0.03)	0.27 (0.03)
b	MRL	0.66 (0.04)	0.63 (0.03)	0.60 (0.04)	0.55 (0.04)	0.57 (0.04)	0.46 (0.03)	0.44 (0.04)	0.40 (0.04)
b	NNU	0.73 (0.03)	0.64 (0.03)	0.62 (0.03)	0.57 (0.04)	0.55 (0.04)	0.44 (0.03)	0.29 (0.03)	0.25 (0.03)
b	NNL	0.70 (0.04)	0.65 (0.03)	0.65 (0.04)	0.61 (0.03)	0.60 (0.04)	0.50 (0.03)	0.45 (0.03)	0.38 (0.03)
MAE	MRU	8.24 (0.43)	12.18 (0.60)	10.07 (0.50)	12.95 (0.69)	11.36 (0.53)	13.46 (0.61)	8.71 (0.35)	1.89 (0.56)
MAE	MRL	7.69 (0.38)	11.06 (0.54)	9.01 (0.49)	12.04 (0.69)	10.26 (0.54)	12.48 (0.62)	7.20 (0.31)	10.78 (0.46)
MAE	NNU	7.01 (0.34)	11.06 (0.55)	9.30 (0.44)	12.36 (0.62)	10.93 (0.53)	12.41 (0.54)	7.92 (0.35)	11.79 (0.53)
MAE	NNL	7.06 (0.34)	10.42 (0.55)	8.57 (0.46)	11.72 (0.62)	10.01 (0.53)	11.79 (0.55)	7.08 (0.30)	10.81 (0.47)
RMSE	MRU	10.71 (0.63)	15.89 (0.83)	12.99 (0.74)	17.53 (1.04)	14.46 (0.76)	16.60 (0.78)	10.65 (0.39)	15.00 (0.70)
RMSE	MRL	10.02 (0.56)	14.73 (0.83)	12.04 (0.74)	16.48 (0.97)	13.44 (0.79)	15.55 (0.72)	8.82 (0.35)	13.55 (0.59)
RMSE	NNU	9.04 (0.55)	14.41 (0.81)	11.99 (0.60)	16.41 (0.88)	13.95 (0.76)	15.37 (0.68)	9.84 (0.40)	14.85 (0.69)
RMSE	NNL	9.24 (0.54)	13.98 (0.84)	11.45 (0.67)	15.57 (0.89)	13.12 (0.81)	14.67 (0.64)	8.64 (0.35)	13.55 (0.61)
$RMSE_s$	MRU	5.99 (0.73)	10.39 (1.07)	8.96 (1.03)	13.24 (1.33)	10.69 (0.90)	12.87 (0.86)	9.47 (0.45)	12.44 (0.81)
$RMSE_s$	MRL	5.70 (0.71)	9.07 (1.01)	7.76 (1.02)	11.58 (1.23)	9.04 (0.88)	11.35 (0.85)	6.48 (0.45)	10.29 (0.75)
$RMSE_s$	NNU	4.52 (0.55)	8.81 (0.89)	7.33 (0.83)	11.01 (1.20)	9.53 (0.89)	11.76 (0.78)	8.24 (0.45)	12.71 (0.77)
$RMSE_s$	NNL	5.07 (0.62)	8.56 (0.90)	6.81 (0.93)	9.99 (1.16)	8.34 (0.89)	10.67 (0.76)	6.43 (0.40)	10.59 (0.74)
$RMSE_u$	MRU	8.86 (0.51)	11.99 (0.62)	9.36 (0.46)	11.44 (0.64)	9.71 (0.58)	10.45 (0.67)	4.86 (0.25)	8.36 (0.43)
$RMSE_u$	MRL	8.22 (0.42)	11.58 (0.60)	9.17 (0.45)	11.67 (0.68)	9.92 (0.53)	10.61 (0.59)	5.97 (0.32)	8.80 (0.38)
$RMSE_u$	NNU	7.82 (0.48)	11.38 (0.67)	9.46 (0.49)	12.13 (0.57)	10.17 (0.49)	9.87 (0.52)	5.37 (0.26)	7.66 (0.30)
$RMSE_u$	NNL	7.70 (0.42)	11.03 (0.60)	9.17 (0.47)	11.90 (0.69)	10.11 (0.51)	10.04 (0.52)	5.76 (0.23)	8.43 (0.35)
d_1	MRU	0.66 (0.02)	0.64 (0.02)	0.61 (0.02)	0.60 (0.02)	0.60 (0.02)	0.53 (0.02)	0.35 (0.03)	0.42 (0.02)
d_1	MRL	0.68 (0.02)	0.68 (0.02)	0.66 (0.02)	0.64 (0.02)	0.65 (0.02)	0.57 (0.02)	0.53 (0.02)	0.51 (0.02)
d_1	NNU	0.72 (0.02)	0.68 (0.02)	0.65 (0.02)	0.64 (0.02)	0.63 (0.02)	0.57 (0.02)	0.45 (0.03)	0.42 (0.02)
d_1	NNL	0.71 (0.02)	0.70 (0.02)	0.68 (0.02)	0.66 (0.02)	0.67 (0.02)	0.60 (0.02)	0.54 (0.02)	0.51 (0.02)
d_2	MRU	0.86 (0.02)	0.85 (0.02)	0.82 (0.02)	0.81 (0.02)	0.81 (0.02)	0.73 (0.02)	0.54 (0.04)	0.63 (0.03)
d_2	MRL	0.88 (0.01)	0.87 (0.01)	0.86 (0.02)	0.84 (0.01)	0.85 (0.02)	0.79 (0.02)	0.76 (0.03)	0.73 (0.03)
d_2	NNU	0.91 (0.02)	0.88 (0.02)	0.86 (0.01)	0.85 (0.01)	0.84 (0.02)	0.78 (0.02)	0.66 (0.03)	0.62 (0.03)
d_2	NNL	0.90 (0.01)	0.89 (0.01)	0.88 (0.01)	0.87 (0.01)	0.86 (0.02)	0.81 (0.02)	0.77 (0.02)	0.73 (0.03)
R^2	MRU	0.59 (0.05)	0.57 (0.04)	0.54 (0.03)	0.54 (0.03)	0.53 (0.04)	0.38 (0.05)	0.15 (0.04)	0.23 (0.04)
R^2	MRL	0.64 (0.04)	0.63 (0.03)	0.60 (0.03)	0.59 (0.03)	0.59 (0.04)	0.46 (0.04)	0.41 (0.05)	0.37 (0.05)
R^2	NNU	0.70 (0.05)	0.65 (0.04)	0.61 (0.04)	0.59 (0.03)	0.56 (0.04)	0.47 (0.05)	0.27 (0.05)	0.24 (0.04)
R^2	NNL	0.69 (0.04)	0.67 (0.04)	0.64 (0.03)	0.63 (0.03)	0.61 (0.04)	0.52 (0.05)	0.44 (0.05)	0.37 (0.05)

parsimony caused by having all four variables in the regression. Regressions were run on the SYSTAT® software package.

Although there is no single, standard neural network, the multilayer perceptron incorporating backpropagation of error (as outlined earlier and illustrated in Figure 2) is a widely used network architecture for regression-like purposes. Following the above rationale, two appropriate (and typical) network constructions were selected, after experimentation with the number of nodes in the hidden layer and the number of hidden layers (to avoid an oversimplified or unnecessarily complex network structure). It should be noted that the number of nodes in the hidden layer is not directly analogous to the number of coefficients in a regression, and that there is no "correct" choice for the number of nodes (although many rules-of-thumb exist, most lack broad theoretical justification). For neural networks not utilizing the lagged ozone data, a 4 x 6 x 1 structure was used (four nodes in the input layer, six nodes in the hidden layer, and one node for the output layer), while a 5 x 7 x 1 structure was used for networks that included the lagged data as input. The neural networks were run for 10,000 iterations on the training data (at which point marginal errors were negligibly small and highly asymptotic). The neural network modeling was performed using the NevProp software package.[22]

Four combinations of models and data were compared for each city: multiple regression using only the unlagged data (MRU), multiple regression using the lagged data (MRL), neural networks using the unlagged data (NNU), and neural networks using the lagged data (NNL). Depending on the data being compared, the use of single statistics such as R^2 can lead to incomplete (or even misleading) results.[23,24] Therefore, a range of statistics for model evaluation and comparison was used in this study, as recommended and described by several authors.[7,23-26] Bootstrap estimates of central tendency (means) and dispersion (standard errors) were calculated for the various statistics, which provided non-parametric measures of confidence on a case-by-case basis for each statistic.[25,27]

RESULTS

After model development, predictions of daily maximum ozone (P_i) from each of the four models were produced for the 250 observations (O_i) in the test data set for each city. Bootstrap estimates of means and standard errors were calculated using 1,000 replications resampled from these data (each comprising 250 values, chosen randomly with replacement). The resulting distributions of bootstrap samples were very close to normal in all cases. All validation measures discussed below refer to these observed and predicted daily maximum ozone data; the full sets of these evaluation and comparison statistics are reported with their standard errors in Table 2.

Summary Univariate Statistics

The means of the observed ozone data ($\bar{O}$) for the eight cities range between 39.50 ppb at the Chicago monitoring site and 66.49 ppb at the Atlanta site. Predicted ozone concentrations ($\bar{P}$) are quite close to these values, on average, within 1 to 2 ppb of the observed means (Table 2). The difference $\bar{P}$-$\bar{O}$ is the residual, or mean bias error (MBE). The MBE is technically a difference statistic (see the section on error indices below), but it is included here because it succinctly summarizes the average over- or under-prediction of ozone in each city (Figure 4). The MBE values are small, and the sizes of the standard errors indicate no discernible differences between models. Comparing cities, all models tend to slightly over-predict in Pittsburgh and Tucson, for example, and slightly under-predict in Phoenix and Atlanta. These subtle variations in MBE may be attributable to slight differences in information between the training data set and the independent test data set for each city, rather than any overall structural tendency in the models.

The standard deviations of the observations (s_o) range from 11.42 ppb in Tucson to 25.54 ppb in Atlanta. The s_o values generally match expectations of variability (e.g., the southwestern cities have lower s_o values and a less variable weather-ozone regime, and likewise for higher s_o values and higher-variability regimes in the eastern cities). The same general pattern is seen in the standard deviations of the predictions (s_p), but with two specific differences. First, the s_p values are consistently lower for all cities; second, there are notable contrasts between models (Table 2). Figure 5 illustrates s_p - s_o , which highlights these differences. The overall negative values reflect the inability of all four models to accurately reproduce the original variability in the data. This is because empirically derived models such as these produce

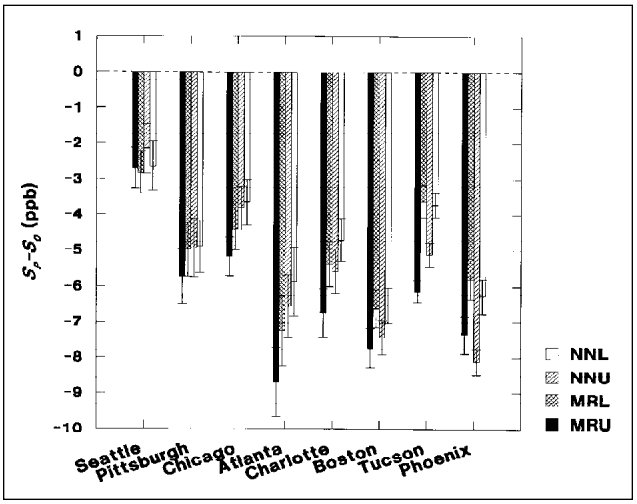

Figure 5. Residuals between the standard deviations of ozone predictions and observations, by model and city. Bootstrap standard error bars are for s_p alone, they are provided as a comparative guide to dispersion.

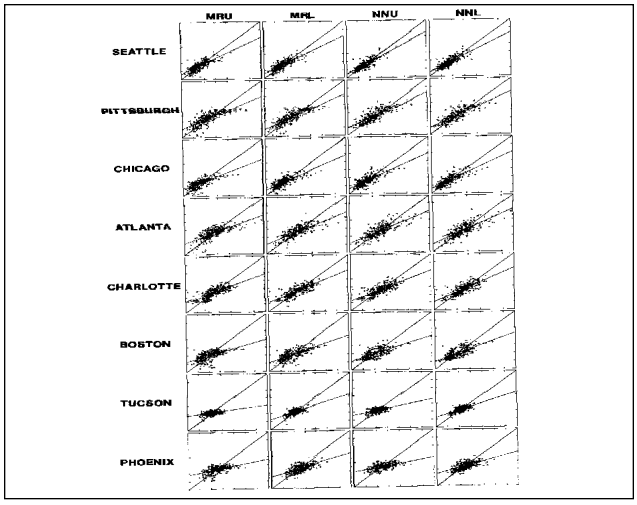

Figure 6. Scatterplots of ozone observations (horizontal axes) against model predictions (vertical axes), for the 250-member test data set. The scale on all axes is from 0 to 175 ppb. The two diagonals in each plot indicate the best-fit regression line through the data and the line of perfect correspondence between observations and predictions.

a best-fit or "average" prediction for a specified set of inputs, whereas a range of observations might, in reality, result from identical inputs (including the tails of the distribution, as discussed earlier). Therefore, models such as these are not capable of exactly reproducing original variability, although a stochastic component or an adjustment procedure can be included.[7] Nonetheless, better models will more closely approximate the original variability. Although distinctions between models are weak (except for Tucson and Phoenix), the overall tendency is for the neural network models to have smaller s_p - s_o values than the regression models. Also, both forms of lagged model (MRL and NNL) have better variability than their unlagged forms (MRU and NNU). The models perform best on the Seattle data, where s_p values are within 2 to 3 ppb of s_o, most likely due to the combination of low $\bar{O}$ and low s_o.

Intercept and Slope

Model predictions that exactly match empirical observations would, in a scatterplot of individual O_i and P_i values, fall on a straight line with an intercept (a) of 0.0 and a slope (b) of 1.0. Thus, in addition to the graphical information on variability and range provided by a scatterplot, a and b quantify useful information on systematic (linear) over- or under-prediction by models.[7,23] Values for a and b by model and city are reported in Table 2. Values for a range between 11.29 ppb for the NNU model in Seattle and 45.69 ppb for the same model in Phoenix. Values for b range from 0.73 for the NNU model in Seattle to 0.18 for the MRU model in Tucson. As expected, the lower intercepts generally coincide with the higher slopes, indicating that the models are overpredicting low values and underpredicting high values (in accordance with previous comments on variability). The NNL models are least susceptible to this tendency, while it is most noticeable for MRU models. For the less well-predicted cities (e.g., Phoenix and Tucson), it appears that the lagged ozone data make an important contribution to regression and neural network models (MRL and NNL), with well-separated standard errors. In contrast, the models are not well-separated in the better-predicted cities (e.g., Seattle), but the broad pattern favors the NNU and NNL models (in terms of a and b) as compared to either of the MRU or MRL models.

To provide a visual reference for interpreting the results in Table 2, the full set of scatterplots illustrating observed versus predicted ozone by city and model are provided in Figure 6, including the best-fit lines described by a and b. The differences in model performance between the cities are easily discernible, but differences within any particular city are not as easy to determine visually. It is quite clear that the models for cities in the upper part of Figure 6 perform better than those in the lower part, in terms of the scatter of data points (s_o and s_p) and the intercept and slope of the best-fit lines (a and b) in each set of graphs. While all models visibly underpredict high-ozone episodes, the better models make reasonable, if somewhat scattered, predictions in the middle and lower ranges of ozone concentrations. Note that the range of predictive power between cities such as Seattle and Phoenix seems unrelated to either the concentration ($\bar{O}$) or the range of observations in the city (with the possible exception of Tucson).

Error Indices

Measures of difference, or error, are based on the residuals between individual predictions and observations (P_i - O_i).

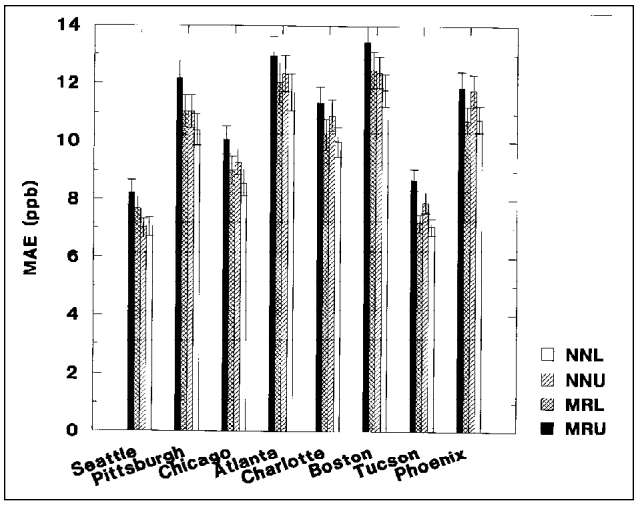

Figure 7. Mean absolute error and associated bootstrap estimates of standard error, by model and city.

The mean absolute error (MAE) is simply the average absolute value of all such deviations, without exponentiation; the root mean square error (RMSE, as in eq 5) is the square root of the mean of all squared residuals (the root is taken to return the result to the original metric which, in this case, is ppb of ozone). MAE has the benefit of being fairly intuitive to interpret, and it is not sensitive to outliers, but RMSE is widely used and is more amenable to additional statistical analyses.[24,26]

MAE ranges between 7.01 ppb for NNU in Seattle and 13.46 ppb for MRU in Boston. Values for MAE in Table 2 are also illustrated in Figure 7, to highlight the pattern of model errors. These errors are somewhat proportional to $\bar{O}$ on a city-by-city basis and, while there is some overlap of standard errors, the relative performance of the models is quite consistent. Generally, neural networks perform better, as do both kinds of models incorporating lagged ozone data. Thus, the NNL model has the overall lowest MAE values and the MRU model the highest. The results for RMSE follow a remarkably similar pattern to MAE, but with slightly higher values because of the outlier sensitivity. The RMSE results in Table 2 show the broad city-based dependency on $\bar{O}$, as well as the better performance of neural networks and both forms of lagged model (lowest RMSE of 8.64 ppb for NNL in Tucson, highest of 17.53 ppb for MRU in Atlanta). Both the MAE and RMSE results provide interesting contrasts in prediction errors between the various cities. Model errors for Tucson are actually as low as those for Seattle, yet the other model evaluation statistics show far poorer model performance for the Tucson data. Tucson has a small MBE and the smallest s_p, with high intercept and low slope values on the scatterplot of O_i against P_i. Therefore, the Tucson results are tightly packed within a small range (see Figure 6), but, in a relative sense, the over- and under-predictions are worse than those for Seattle.

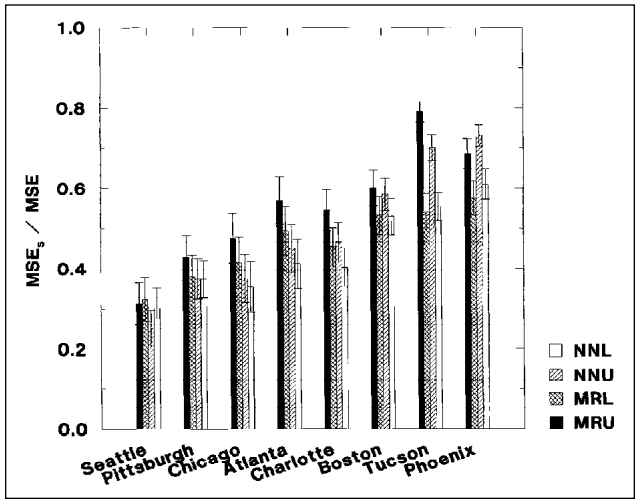

Figure 8. Proportion of systematic (model-oriented) errors and associated, bootstrap estimates of standard error, by model and city.

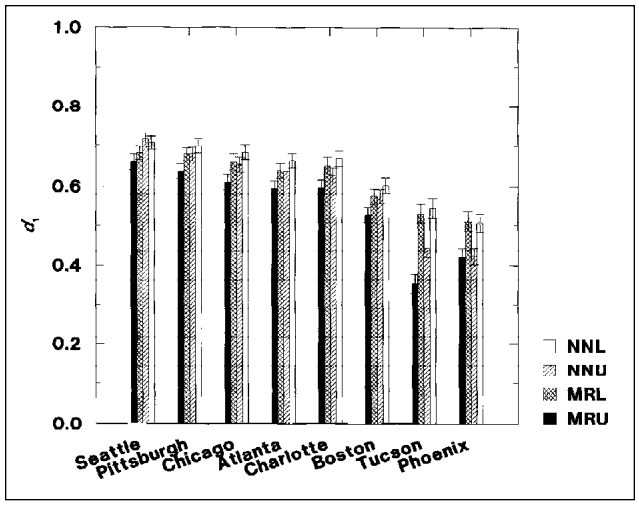

Figure 9. Index of agreement based on simple (unsquared) differences and associated estimates of standard error, by model and city.

Such relative contributions of model- and data-oriented error may be partitioned into the systematic and unsystematic portions of RMSE (RMSE$_s$ and RMSE$_u$). Systematic, or model-oriented, error is based on the differences between expected predictions and observations $\hat{P}_i - O_i$ (where $\hat{P}_i = a + bO_i$), and unsystematic or data-oriented error is based on $P_i - \hat{P}_i$, the differences between actual and expected predictions.[23,25] RMSE$_s$ and RMSE$_u$ are reported in Table 2, and it can be seen that the same patterns of model error emerge as before. That is, while there is some overlap of standard error, neural networks have generally lower systematic errors, as do the models using lagged ozone data. Unsystematic errors are relatively inconsistent between models, but as with RMSE, both forms of error are related to the size of $\bar{O}$ for each city. Partitioning of error is important because it indicates the degree to which individual models can be improved within the constraints of variability in the observed data. Comparisons of the two sources of error are made easier by examining the systematic error as a proportion of the total error, calculated from the squares of the two measures as MSE$_s$/MSE (the proportion of unsystematic error is 1 - MSE$_s$/MSE). Figure 8 illustrates MSE$_s$/MSE, which ranges from 0.25 for NNU in Seattle to 0.79 for MRU in Tucson. Again, standard errors do not permit complete separation of the models in many cases, but generally the MRU and NNU models do not perform as well as the MRL and NNL models, especially in poorly predicted cities such as Tucson and Phoenix (where the persistence information from the lagged data leads to substantial drops in model error).

Measures of Agreement

Perhaps the most common measures of agreement are Pearson's product-moment correlation coefficient (R) and its square, the coefficient of determination (R^2) or explained variance. Unfortunately, the values of these measures might be unrelated to the sizes of P_i - O_i, and therefore might be misleading. Consider two coincident sine waves, one having twice the amplitude of the other: both R and R^2 will equal 1.00, but there are important differences not captured by these measures. To circumvent this correlation problem, Willmott[23,25] advocates an index of agreement (d_α) that measures the degree to which a model's predictions are error free. The value of d_α is expressed as:

$$d_\alpha = 1 - \left[\sum_{i=1}^{n} |P_i - O_i|^\alpha \right] \left[\sum_{i=1}^{n} \left(|P_i - \overline{O}| + |O_i - \overline{O}| \right)^\alpha \right]^{-1} \quad (8)$$

where $\alpha = 1$ or $\alpha = 2$. The d_1 statistic is based on simple differences and is more conservative than d_2, which is based on squared differences. Both versions of d_α indicate the degree to which predicted deviations about $\overline{O}$ differ from observed deviations about $\overline{O}$, considering both sign and magnitude. For the sine wave example above, d_1 = 0.67 and d_2 = 0.89. Both forms of d_α and R^2 were calculated to evaluate the models, d_α because of the above properties, and R^2 because of its widespread use.

The model comparison results for d_1, d_2, and R^2, while different numerically, are practically identical in terms of relative model performance. For completeness, results for all three statistics are presented in Table 2, but those for d_1 are also illustrated in Figure 9 to highlight the relative ranking of each model. Willmott's d_1 ranges between 0.35 for MRU in Tucson and 0.72 for NNU in Seattle, while the matching values for d_2 range from 0.54 and 0.91, and for R^2 from 0.15 to 0.70. Of the eight cities, five (Seattle, Pittsburgh, Chicago, Atlanta, and Charlotte) have moderate to good d_1, d_2, and R^2 values. The results for Tucson and Phoenix are poor, while the Boston results are intermediate. Within the better-predicted cities, the models all range over about 10% explained variance (approximately 0.08 for d_1 and 0.05 for d_2). Also, the poorest and best model performances in any of these cities are well-seperated in terms of standard error, but the intermediate models are not easily differentiated. In Chicago, for example, the improvement between models comprises a clear jump in d_1, from 0.61 (MRU) to 0.68 (NNL), but MRL and NNU are indistinguishable (0.66 ± 0.02 and 0.65 ± 0.02, respectively). Generally, the same patterns of model performance emerge from these statistics as from those examined earlier. The MRU model is uniformly the poorest performer, and the MRL model improves over MRU in each instance (quite dramatically in the poorly predicted cities). The NNU model sometimes improves slightly over MRL, but in the poorly predicted cities NNU performs substantially worse than MRL. The NNL model is the best in all but the Seattle case, and while this is a consistent pattern, it is not well-separated from the intermediate MRL or NNU models. It is unclear why NNL is marginally less successful than NNU for the Seattle data, but in terms of standard error, there is no definitive difference between either neural network model.

CONCLUSIONS

Overall, it appears that neural networks are somewhat, but not overwhelmingly, better than multiple regression models for weather-based ozone forecasting. Perhaps the literature on neural networks encourages high performance expectations, but this study found no dramatic improvements. On the other hand, this study also showed that neural networks are consistently better than regression models for ozone forecasting, although the gains in performance are only small to moderate. It would seem that the inherent incorporation of nonlinear relationships into neural network models allows them to make somewhat more accurate predictions of ozone than regression models using the same set of input data. The inclusion of lagged ozone data (the ozone maximum from the previous day) generally improves both kinds of model. It is interesting that the relative improvement between unlagged and lagged models is greater for regression models than neural networks (i.e., neural networks using lagged data do not have to rely as much on persistence information in those data). Also, all models underestimate high-ozone episodes and overestimate low-ozone events. Neural networks improve, but do not remove, this common problem among empirical-statistical models. Generally, the best model is a neural network incorporating lagged ozone data, which provides some improvement in performance over both multiple regression models and the unlagged neural network model.

The best-predicted of the eight cities are (in order) Seattle, Pittsburgh, Chicago, Atlanta, and Charlotte. Model performance is very similar for each of these, although this group included the lowest and highest $\overline{O}$. By extension, weather-ozone relationships in these cities are of similar importance (although individual interactions between weather variables and ozone will differ, of course). The differences between the various modeling approaches are most dramatic in the poorly predicted cities (Tucson and Phoenix), especially the improvements resulting from the lagged ozone data. Both cities have prolonged high-ozone conditions (hot, sunny, and frequently less windy) compared to the more changeable weather elsewhere, so the nature of the weather-ozone relationship as it relates to daily ozone variability is far weaker than in other cities. The importance of lagged data for modeling ozone in these cases is

therefore twofold: first, the persistence information is of greater relative significance compared to the weather information; and, second, the persistence information is a better predictor precisely because ozone concentrations change less from day to day. It is unclear why the model-performance results for Boston fell a little below the other better-predicted cities. Certainly, summer weather conditions in Boston are at least as changeable as those in other cities. It is possible that the reasons for the poorer results in Boston are site specific.

The results of this study suggest that weather-ozone relationships might be sufficiently complex that a level of random variation or noise exists (as assumed in some ozone trend studies),[28,29] which cannot be captured by even a relatively sophisticated empirical-statistical model such as a neural network. Inputs of greater sophistication are probably required to extract any remaining non-random unexplained variance in the data. Also, neural networks will appear to some as more complex or less familiar than traditional regression modeling. Results using multiple regression with lagged ozone data were almost as good as with the corresponding neural network models, and it could be argued that regression models constitute a reasonable, practical choice in cases where absolute predictive power is not essential. This study focuses on making directly comparable validations of the various models to identify the best general modeling approach to use. In an operational setting for any specific city, individual regression and neural network models could be improved through experimentation and fine-tuning. Likely strategies include using additional variables (e.g., extra weather elements, synoptic patterns, traffic flow or day-of-the-week information) or changing the nature of the model (e.g., incorporating a priori curvilinear transformations of input data for regressions, or changing the number of hidden nodes and layers in neural networks).

REFERENCES

1. Seinfeld, J.H. *Atmospheric Chemistry and Physics of Air Pollution*; Wiley: New York, 1986.
2. Boubel, R.W.; Fox, D.L.; Turner, D.B.; Stern, A.C. *Fundamentals of Air Pollution*; Academic Press: San Diego, 1994.
3. Godish, T. *Air Quality*; Lewis: Chelsea, MI, 1991.
4. Wolff, G.T.; Lioy, P.J. "An empirical model for forecasting maximum daily ozone levels in the northeastern United States," *Atmos. Environ.* **1978**, *11*, 967-983.
5. Clark, T.L.; Karl, T.R. "Application of prognostic meteorological variables to forecasts of daily maximum one-hour ozone concentrations in the northeastern United States," *J. Applied Meteorology* **1982**, *21*, 1662-1671.
6. Robeson, S.M.; Steyn, D.G. "A conditional probability density function for forecasting ozone air quality data," *Atmos. Environ.* **1989**, *23*, 689-692.
7. Robeson, S.M.; Steyn, D.G. "Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations," *Atmos. Environ.* **1990**, *24B*, 303-312.
8. Feister, U.; Balzer, K. "Surface ozone and meteorological predictors on a subregional scale," *Atmos. Environ.* **1991**, *25A*, 1791-1790.
9. Eder, B.K.; Davis, J.M.; Bloomfield, P. "An automated classification scheme designed to better elucidate the dependence of ozone on meteorology," *J. Applied Meteorology* **1994**, *33*, 1182-1199.
10. Burrows, W.R.; Benjamin, M.; Beauchamp, S.; Lord, E.R.; McCollor, D.; Thomson, B. "CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal and Atlantic regions of Canada," *J. Applied Meteorology* **1995**, *34*, 1848-1862.
11. Ryan, W.F. "Forecasting severe ozone episodes in the Baltimore metropolitan area," *Atmos. Environ.* **1995**, *29*, 2387-2398.
12. Hewitson, B.C.; Crane, R.G. *Neural Nets: Applications in Geography*; Kluwer: Boston, 1994.
13. Collins, E; Ghosh, S.; Scofield, S. *Risk Analysis: DARPA Neural Network Study*; AFCEA International Press: Fairfax, VA, 1988.
14. Marzban, C.; Stumpf, G.J. "A neural network for tornado prediction based on Doppler radar-derived attributes," *J. Applied Meteorology* **1996**, *35*, 617-626.
15. Boznar, M.; Lesjak, M.; Mlakar, P. "A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain," *Atmos. Environ.* **1993**, *27B*, 221-230.
16. Ruiz-Suárez, J.C.; Mayora-Ibarra, O.A.; Torres-Jiménez, J; Ruiz-Suárez, L.G. "Short-term ozone forecasting by artificial neural networks," *Advances Engineering Software* **1995**, *23*, 143-149.
17. Rumelhart, D.E.; McClelland, J.L. *Parallel Distributed Processing*, Vols. 1 and 2; MIT Press: Cambridge, MA, 1986.
18. Anderson, J.A.; Rosenfeld, E. *Neurocomputing: Foundations of Research*; MIT Press: Cambridge, MA, 1988.
19. Comrie, A.C. "The climatology of surface ozone in rural areas: a conceptual model,"*Progress in Physical Geography* **1990**, *14*, 295-316.
20. Comrie, A.C.; Yarnal, B. "Relationships between synoptic-scale atmospheric circulation and ozone concentrations in metropolitan Pittsburgh, Pennsylvania," *Atmos. Environ.* **1992**, *26B*, 301-312.
21. Comrie, A.C. "An all-season synoptic climatology of air pollution in the U.S.-Mexico border region," *Professional Geographer* **1996**, *48*(3), 237-251.
22. Goodman, P.H. *NevProp Software*; Center for Biomedical Modeling Research: University of Nevada, Reno (URL ftp://ftp.scs.unr.edu/pub/cbmr/nevpropdir/).
23. Willmott, C.J. "On the validation of models," *Physical Geography* **1981**, *2*, 184-194.
24. Willmott, C.J. "Some comments on the evaluation of model performance," *Bulletin American Meteorology Soc.* **1982**, *63*, 1309-1313.
25. Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O'Donnell, J.; Rowe, C.M. "Statistics for the evaluation and comparison of models," *J. Geophysical Research* **1985**, *90*(C5), 8995-9005.
26. Fox, D.G. "Judging air quality model performance: A summary of the AMS Workshop on Dispersion Model Performance," *Bulletin American Meteorology Soc.* **1981**, *62*, 599-609.
27. Efron, B.; Gong, G. "A leisurely look at the bootstrap, the jacknife, and cross-validation," *American Statistician* **1983**, *37*, 36-48.
28. Rao, S.T.; Zurbenko, I.G. "Detecting and tracking changes in ozone air quality," *J. Air Waste Manage. Assoc.* **1994**, *44*, 1089-1092.
29. Flaum, J. B.; Rao, S.T.; Zurbenko, I.G. "Moderating the influence of meteorological conditions on ambient ozone concentrations," *J. Air Waste Manage. Assoc.* **1996**, *46*, 35-46.

About the Author

Dr. Andrew C. Comrie is an assistant professor in the Department of Geography and Regional Development, University of Arizona, Harvill Building Box #2, Tucson, AZ 85721. His research specialties include climatology and air quality. Send correspondence to the above address or e-mail to comrie@geog.arizona.edu.