Pergamon

Atmospheric Environment Vol. 29, No. 17, pp. 2387–2398, 1995

1352–2310/95 $9.50 + 0.00

1352–2310(94)00302–5

FORECASTING SEVERE OZONE EPISODES IN THE BALTIMORE METROPOLITAN AREA

WILLIAM F. RYAN
Department of Meteorology, University of Maryland, College Park, Maryland, U.S.A.

(*First received* 14 *April* 1994 *and in final form* 15 *September* 1994)

Abstract—In order to protect sensitive individuals and support possible episodic emissions control efforts, a pilot program was undertaken to forecast episodes of high ozone concentrations for 24–72 h in the Baltimore metropolitan area. This program utilized Classification and Regression Tree (CART) algorithms as well as standard regression analysis and expert forecasts. All approaches tend to underpredict peak O_3 concentrations at 24 h. The low bias varied from 5 ppbv for the expert forecast to 7 ppbv for the CART forecast. The standard error for forecast O_3 varied from 17 ppbv for the regression forecast to 23 for the CART forecast. For high O_3 (> 120 ppbv) events, the expert forecast has the best success with a detection rate of 50% and skill scores varying from 0.40 to 0.50. For expert forecasts of greater than 115 ppbv, verifying against observations of greater than 120 ppbv, the detection rate rises to 75% with skill scores of 0.60–0.66. One of the principal sources of forecast error was underprediction of mid-day surface temperature by the standard meteorological models during extremely warm episodes. When perfect forecasts of meteorological parameters are utilized, the forecast skill of the objective measures increases to approximately the skill of the expert forecasts. Improvements can be made by utilizing more recent data for initialization of the ozone regression equations and the use of local forecasts to supplement temperature forecasts during warm periods.

Key word index: Photochemical ozone, ozone forecasting, CART, Baltimore–Washington.

1. INTRODUCTION

The Baltimore metropolitan area has a serious photochemical ozone (O_3) air pollution problem. The National Ambient Air Quality Standard (NAAQS) for O_3 is 120 parts per billion by volume (ppbv) for a one-hour average. To comply with Clean Air Act (CAA) standards for O_3, the highest daily ozone concentration measured by a monitor in a metropolitan area may not exceed the NAAQS on more than four occasions over a three-year period. For the 11-year period from 1983 to 1993, ozone monitors in the Baltimore metropolitan area measured O_3 concentrations ([O_3]) at or in excess of 120 ppbv on 222 days (Fig. 1). Even the "cleanest" three-year period (1989–1991) includes 47 days in which [O_3] reached or exceeded 120 ppbv.

Despite years of pollution control efforts, this pattern of continued air quality violations is found in a number of other major metropolitan areas (Schulze, 1993; National Research Council, 1991). In response to the apparent intractability of the urban air pollution problem, Congress passed the Clean Air Act Amendments in 1990. These amendments mandate, among other things, enhanced photochemical modeling efforts to identify more effective pollution control strategies. The possibility exists, however, that further large reductions in O_3-precursor emissions will not reduce ambient [O_3] below the NAAQS in all cases. To ensure that the standard is rarely violated, it may be more efficient to impose more sweeping emissions controls only during high ozone episodes. The usefulness of these "episodic" control measures will necessarily depend on effective forecasting of high O_3 episodes. This paper describes a pilot O_3 forecasting program carried out in the Baltimore area in the summer of 1993.

2. REGIONAL OZONE CLIMATOLOGY AND CART DEMONSTRATION

Photochemical air pollution, of which O_3 is the principal constituent, is formed by a series of reactions involving hydrocarbons, reactive nitrogen species and sunlight (Seinfeld, 1989). Hydrocarbons are emitted from both natural and anthropogenic sources. The latter include mobile sources (automobiles) and stationary sources (industrial uses). Reactive nitrogen (NO_x) is emitted mainly by automobiles and large combustion sources such as power plants. Large metropolitan areas, such as Baltimore, typically have a large commuting population and hence large mobile source emissions of hydrocarbons and NO_x. In addition, the power requirements of a large city necessitate many local power plants. Regional emissions of O_3-precursors can also effect local air quality. Baltimore is situated within 200 km of two major metropolitan areas, Philadelphia and Washington, D.C. (Fig. 2).

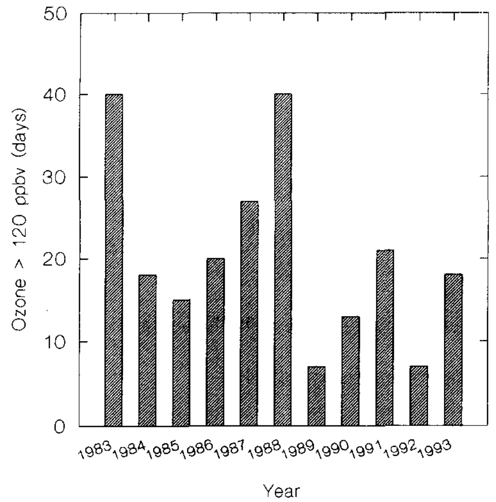

Fig. 1. Number of days in which O_3 monitors in the Baltimore metropolitan area observed [O_3] in excess of 120 ppbv (one hour average) for the period 1983–1993.

Both of these cities emit significant amounts of O_3-precursors and experience high ambient [O_3]. To the west of Baltimore, typically the upwind direction in the prevailing westerlies, is the industrialized Midwest and Ohio Valley with significant elevated sources of NO_x. Emissions from these regions can be transported to the Baltimore area and affect local air quality (Moy *et al.*, 1994). Any forecasting program must take into account both local conditions conducive to O_3 production (e.g. temperature, sky cover, surface winds) as well as regional conditions that may effect local pollution levels (e.g. upper air wind speed and direction).

The O_3 problem in the Baltimore region is a summer phenomenon associated with persistent weather patterns. For those ozone-monitoring stations within 50 km of the Baltimore urban center during the period 1983–1993, 88% of days with observed [O_3] greater than 120 ppbv occur in the months of June through August (Fig. 3). If the last three days in May are included, the percentage rises to 92%. A similar analysis for the Baltimore–Washington region, including northern Virginia, for the period 1983–1990, showed that 85% of all occurrences of monitored [O_3] in excess of 120 ppbv occurred in the June–August period.

The association of O_3 episodes with persistent weather patterns is suggested by the frequency of multi-day O_3 episodes. For the Baltimore area, 159 of 227 ozone violation days occurred sequentially. The more severe events, characterized by three or more monitors observing [O_3] in excess of 120 ppbv, are more likely to be multi-day events with 82% occurring in groupings of sequential days. A similar frequency (83%) was found for severe events in the larger Baltimore–Washington region, with severe events here characterized by five or more monitors in violation. In addition, for the period 1987–1990, 19 of the 20 most severe cases in the Baltimore–Washington region were part of multi-day events.

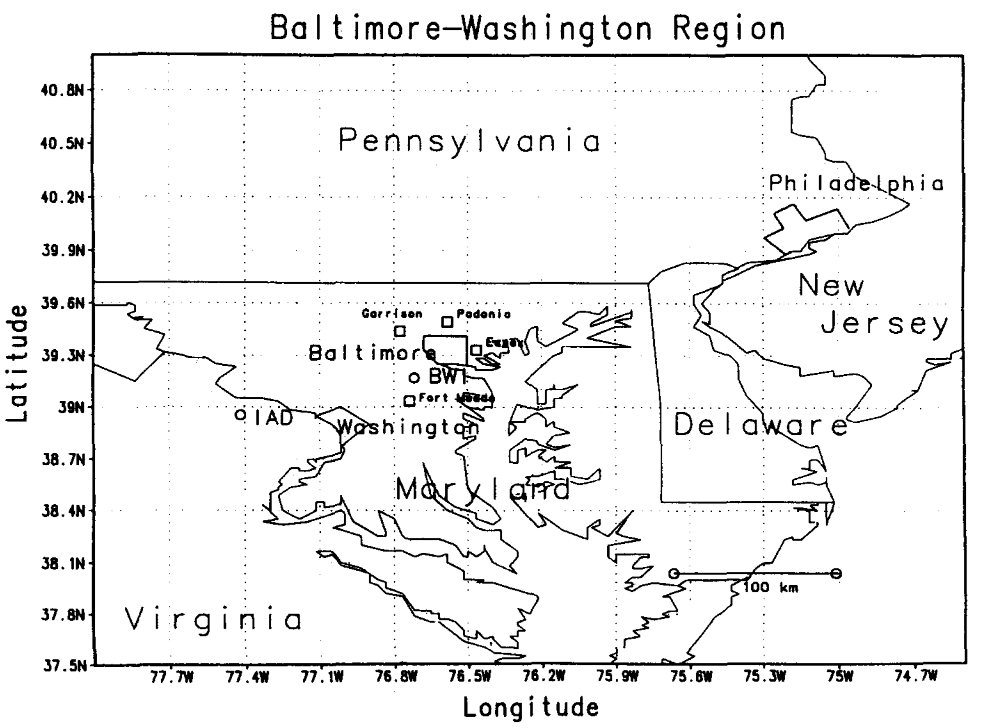

Fig. 2. Map of the Baltimore–Washington region.

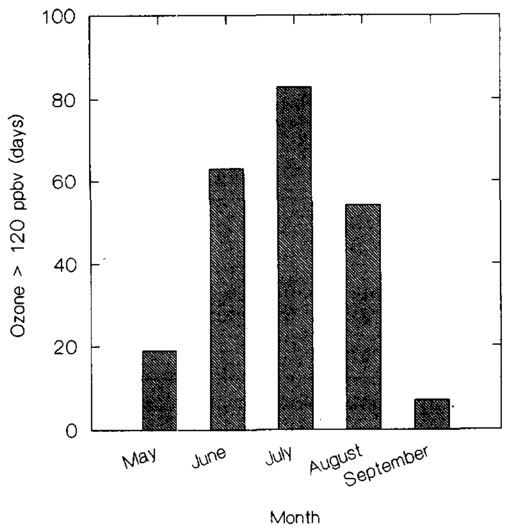

Fig. 3. Monthly distribution of number of days in which O_3 monitors in the Baltimore metropolitan area observed [O_3] in excess of 120 ppbv (one hour average) for the period 1983–1993.

The forecast problem for Baltimore is to identify weather patterns characteristic of strong ozone events and associate these patterns with standard meteorological variables. Previous studies have associated ozone episodes along the East Coast with the slow passage of anticyclones (Vukovich *et al.*, 1977; Vukovich and Fishman, 1986). Weather factors associated with anticyclonic conditions provide many of the requirements for photochemical ozone production: clear skies increase photolytic activity, light winds keep local emissions from being diluted, and subsidence at upper levels restricts the depth of the turbulent layer into which emissions are mixed. However, the position, movement and strength of a particular anticyclone can introduce significant variability in the factors noted above. For example, anticyclones associated with outbreaks of cool Canadian air are characterized by strong subsidence but also with cooler temperatures and high winds which will reduce the production of O_3. Conversely, anticyclones associated with the "Bermuda High" may bring high temperatures but also tend to advect cleaner maritime air into the region.

In order to isolate ozone episodes into coherent groups with standard meteorological variables, an objective analysis using the Classification and Regression Tree (CART) technique is utilized. The CART analysis (California Statistical Software, 1991) has been used for general weather forecasting (Burrows, 1991) and has been recommended for use in O_3 studies (Horie, 1987; Seinfeld, 1988; EPA, 1991c; National Research Council, 1991). The CART technique operates by a binary splitting of data into groups that are more homogeneous compared to non-group members (Brieman *et al.*, 1984). A succession of binary splits results in a "tree" whose "branches", or terminal nodes, represent distinct classes, or categories, of data. CART is non-parametric, that is, no assumptions are made regarding the statistical distribution of the predictors. For forecasting purposes, the CART approach is valuable because it is easy to use and allows for a categorical forecast based on a few basic measures. The shortcoming of this approach is the possibility of misclassification which can result in consistent overprediction or underprediction.

The CART analysis begins with a learning sample consisting of a predictand values and predictor values. It then makes a "best" decision tree that separates events into categories of the predictand based on predictor values. The optimum number of categories, or terminal nodes, is that set which results in the minimum "cost" of miscalculation. This "cost" value is determined by holding out a portion of the data set for insertion as test cases. For small data sets, as in this study, the testing process can be run iteratively using a rotating series of test samples. In this study, a tenfold iteration, or cross-validation, was utilized. In the initial runs, the predictand is the domain mean maximum O_3 for the B/W domain for the period June–August 1983–1990. The predictors consist of a primary set of surface observations from Baltimore–Washington International Airport (BWI) and upper air observations at Dulles International Airport (IAD) (Fig. 2). The surface variables are pressure, temperature (T), dew point temperature (T_d), wind velocity and total opaque sky cover for 1000, 1400, 1800 and 2200 UTC (the conversion from UTC to EDT is 4 h so that 1800 UTC is 1400 EDT). The upper air variables are T, T_d, wind velocity and geopotential height (Z) for 50 mb intervals between 950 and 650 mb. In addition, virtual potential temperature (θ_v) is derived for each 50 mb level and gradients for θ_v calculated.

There are a number of shortcomings associated with the data base utilized by the CART analysis. Because the CART analysis is based on meteorological data from one location it cannot fully characterize the regional transport patterns associated with high ozone in the Baltimore–Washington region. The CART analysis may also be limited by changes in the emissions base from 1983 to 1990. This period saw a considerable change in some emissions factors, principally reductions in mobile source emissions from new model cars, though the overall trend appears to be flat (EPA, 1991a). This period also includes the extreme drought summer of 1988 and the anomalous weather conditions associated with it (Trenberth *et al.*, 1988).

The initial CART run was able to resolve the most and least severe O_3 cases (Table 1). The dominant predictor is mid-day (1800 UTC) temperature with lesser weight, or fine tuning between terminal nodes, accounted for by sky cover, 950 mb temperature, and morning (1000 UTC) sea level pressure and temperature. The main split between terminal nodes 1–3

Table 1. Classification and Regression Tree (CART) results for the Baltimore–Washington (B/W) region for the period June–August 1983–1990

Terminal node	Number of cases	Median $[O_3]$ (ppbv)	$[O_3]$ > 120 ppbv (*N*)	$[O_3]$ > 120 ppb (%)	$[O_3]$ > 120 ppbv @ > 5 monitors (*N*)	BWI mean temp (1800 UTC)	BWI mean sky cover (1800 UTC)
1	211	59	4	2	0	78.2	7.0
2	69	43	0	0	0	71.1	9.8
3	179	75	41	23	7	82.4	3.3
4	168	79	50	30	13	88.3	4.3
5	109	100	85	78	41	92.8	2.2
Total	736		180		61		

Concentrations of ozone ($[O_3]$) are given in parts per billion by volume (ppbv) for a one hour average. All surface meteorological data in this and succeeding tables are from the National Weather Service (NWS) station at Baltimore–Washington International Airport (BWI). Observation time for all meteorological data are in Universal Coordinated Time (UTC). The conversion to local time (EDT) is four hours so that 1800 UTC is 1400 EDT.

(TNI–TN3) and TN4–TN5 is based on afternoon temperatures greater than 85.5°F. The strongest node (TN5) is distinguished by higher afternoon temperatures, lower morning temperatures and higher morning pressure. The lower morning temperatures may be reflective of overnight clear skies and strong early morning photolytic activity. This node (TN5) has a strong (78%) probability of containing $[O_3]$ in excess of 120 ppbv and also accounts for 67% of the most severe ozone cases when severity is defined by five or more domain monitors in excess of 120 ppbv. When the most severe ozone episodes are ranked by terminal node, 28 of 34 ozone days are classified in TN5. On the other side of the spectrum, TN1 and TN2 are clearly non-ozone days with cool and cloudy conditions.

While the CART analysis is able to differentiate well between severe ozone cases and non-ozone cases on a regional basis with good ability, there are a large number of strong ozone days associated with the "middle" nodes (TN3 and TN4). TN3 contains the "cooler high ozone" cases. The afternoon temperature is relatively low yet low sky cover and warm 950 mb temperatures suggests stable conditions with strong photolytic activity. Although the frequency of severe events is slight in this node (4%), the chance of an O_3 monitor measuring in excess of 120 ppbv is about one in four and raises a problem of underprediction. The remaining node (TN4) can be characterized as a "warmer low ozone" group. The temperatures associated with this group are quite warm yet sky cover and morning temperatures are higher than the stronger nodes which may reflect a reduction of insolation in the morning hours.

The relationship between high $[O_3]$ and mid-day temperature shown in the CART analysis is reflected in the Baltimore–Washington ozone statistics for 1983–1990 (Fig. 4). Note, however, the large variance at higher temperature levels. In some cases, high temperature does not correspond to high $[O_3]$. A similar distribution is seen in monthly statistics for the Baltimore metropolitan area (Fig. 5). A suggestion of a temporal trend in $[O_3]$ is seen in the linear fit for the most recent years (1989–1993). This difference remains even when the anomalous 1988 data are removed.

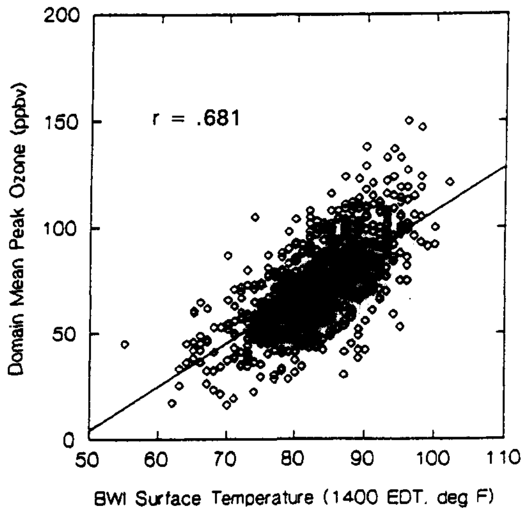

Fig. 4. Plot of domain mean peak $[O_3]$ for the Baltimore–Washington region against mid-afternoon surface temperature at Baltimore–Washington International Airport (BWI) for the period June–August 1983–1990. Straight line represents best linear fit with correlation coefficient of 0.681.

A second series of CART runs investigated those days with peak $[O_3]$ greater than 120 ppbv or domain mean maximum $[O_3]$ greater than 90 ppbv (Table 2). The most severe cases are grouped in TN1 which is distinguished from the rest of the nodes by total θ_v gradient from 950 to 650 mb. TN1 is characterized by high θ_v at 950 mb but a much smaller vertical gradient so that θ_v for all nodes is nearly equal by 750 mb. TN1 may represent "hot" cases in which high temperature and low sky cover drive local O_3 chemistry while TN2–TN4 represent cases in which

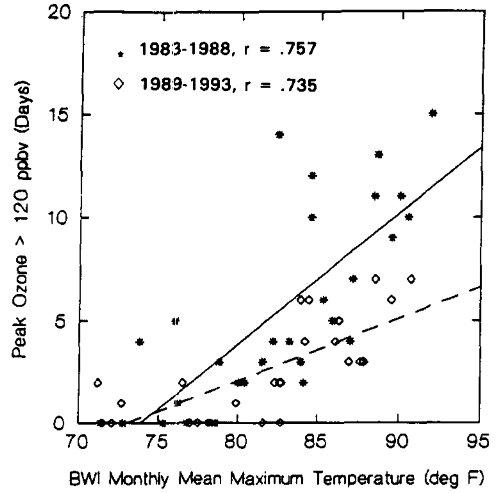

Fig. 5. Plot of monthly mean maximum temperature at BWI against number of days in which O_3 monitors in the Baltimore metropolitan area observed $[O_3]$ in excess of 120 ppbv (one hour average) for the period 1983–1993. Asterisks refer to the period 1983–1988 and diamonds to the period 1989–1993. Best linear fits are shown for each group of years.

stable lower tropospheric conditions trap emissions into a smaller volume. TN3 is characterized by relatively low ozone and no severe cases. This node contains strong westerly winds at 850 mb ($> 5\ \mathrm{m\,s^{-1}}$) as well as strong surface winds. TN3 contains fewer severe cases (17% compared to greater than 50% for TN1 and TN2) and is differentiated by increased cloud cover. For all nodes, the mean v-component of the 850 mb winds was northerly suggesting that high O_3 may be associated with winds having a west or northwest component at the top of the boundary layer.

In order to address the problem of regional transport, a supplementary subjective analysis was undertaken to identify synoptic scale patterns associated with multi-day ozone episodes. This analysis also provides a means for expert augmentation of CART based forecasts. The subjective approach investigated all multi-day ozone episodes for the period June–August 1983–1990 as well as all single day events that contained severe ozone violations (greater than or equal to 5 regional monitors in excess of 120 ppbv). National Weather Service (NWS) surface synoptic charts were analyzed for the position of the nearest sea level pressure maximum (anticyclone center) at 1200 UTC (0800 EDT) for each day of a multi-day event. Local transport was inferred from the orientation of isobars and station wind reports. In addition, the position of nearby frontal zones was noted. The winds at the top of the boundary layer, indicative of longer range transport, are determined by analysis of 850 mb constant pressure upper air charts with the location of the nearest ridge or closed circulation noted.

The subjective analysis identifies four synoptic scale flow patterns associated with multi-day ozone episodes (Table 3). The success rate of this scheme, which measures the degree to which days in a particular episode could be assigned to a particular class, is 86%. The two most common classes reflect the wind results from the CART analyses. Class 1 is characterized by surface high pressure centered south and west of Baltimore, typically in western Virginia or West Virginia with surface winds generally light and westerly. The wind flow at 850 mb is northerly associated with a strong 850 mb ridge or closed circulation. This pattern often occurs with anticyclones that are stationary or retrograding from east to west and was common during the drought summer of 1988. Class 2 is characterized by surface conditions similar to Class 1 in terms of anticyclone position but has a more westerly component at 850 mb. This class is often associated with a stationary front draped east to west across Pennsylvania. In addition, a lee trough is often present in or near the domain in Classes 1 and 2. A lee trough is an area of low pressure located downwind of a mountain range (Pierrehumbert, 1986). This effect is most common in the region east of the Rockies but has been identified as a frequent characteristic of high ozone episodes in the northeastern United States (Pagnotti, 1987). The association of lee troughs with high O_3 may be due to convergence along a trough axis that is often oriented parallel to

Table 2. CART results for the B/W region for the period June–August 1983–1990 for a subset of cases in which O_3 monitors observed $[O_3]$ in excess or 120 ppbv (one hour average) or domain mean peak $[O_3]$ exceeded 90 ppbv

Terminal nodes	Number of cases	Median $[O_3]$	Severe cases	IAD θ_v (950 mb, 1200 UTC)	IAD θ_v (759 mb, 1200 UTC)	BWI sky cover (1800 UTC)	IAD U-wind 850 mb ($\mathrm{m\,s^{-1}}$, 1200 UTC)
1	43	111	27	304.5	308.0	2.8	4.1
2	38	104	17	300.7	308.5	0.4	1.7
3	13	93	0	301.1	309.4	0.5	8.7
4	101	94	17	301.7	308.3	4.2	3.5
Total	192		61				

Upper air data in this and all succeeding tables are from the 1200 UTC radiosonde sounding at Dulles International Airport (IAD). "Severe" ozone cases in this context refers to days in which five or more ozone monitors in the B/W region measured $[O_3]$ in excess of 120 ppbv. The meteorological data are presented as mean data for each group.

Table 3. Results of subjective analysis of multi-day ozone episodes in the B/W for the period May–September 1983–1990

Class	Days with $[O_3] > 120$ ppbv	Monitors with $[O_3] > 120$ ppbv	Brief description of class
1	39	103	Surface high pressure to west; 850 mb winds from north or northeast
2	28	101	Surface high pressure to west; 850 mb winds from west or northwest
3	27	64	"Bermuda High", winds south or southwest
4	25	54	Frontal zone

A brief description of the synoptic weather patterns associated with each class of episodes is given. More specific information is found in the accompanying text.

the northeastern urban corridor. The lee trough is often not well resolved on the synoptic scale and its presence is not an effective tool for discriminating between types of episodes in this type of analysis but its frequent presence was noted during multi-day $[O_3]$ events.

The remaining two classes are less frequent. Class 3 is characterized by high pressure centered to the south and east of Baltimore and often occurs at the end of Class 1 or 2 episodes as the surface anticyclone moves offshore. The "Bermuda High" pattern drives southerly surface winds and is often associated with a frontal zone approaching from the west. This pattern is the most common summertime pattern in this region (NAVAIR, 1966; ESSA, 1968) and may also be associated with high ozone episodes in the northeastern states (EPA, 1991b). A number of ozone episodes occurred with frontal zones at or near the region (Class 4). The association of frontal zones with high ozone poses a problem for any objective model of local ozone. The frontal zone is, by definition, characterized by strong spatial and temporal gradients in temperature, moisture and wind which are not well resolved by point data. In addition, the existence of a frontal zone can provide conditions that are, or are not, conducive to high ozone. For example, convergence along and ahead of frontal zones may well increase local concentrations of ozone precursors. However, deep convection is often likely to form along lines of strong convergence which will tend to vent these precursors and introduce more cloud cover and possibly wet scavenging by precipitation. Regional transport patterns can also be confusing in the vicinity of frontal zones.

Little correlation is found between the subjective classifications and the CART terminal nodes for the subset of severe ozone cases. This is due to several factors. The CART nodes are based on point data and the splits between groups are driven mainly by temperature observations which reflect local ozone production. The subjective classifications, on the other hand, are based on synoptic scale pressure and height patterns. These patterns reflect both local conditions and transport into the region. Unless there is a single dominant temperature and transport pattern, the two approaches will not give identical results. In the Baltimore–Washington region, surface temperature is not necessarily related to synoptic scale weather patterns. For example, hot conditions can be associated with southerly winds, the Bermuda High circulation, and can also be associated with west or northwest winds when the surface ridge is located west of the region. In terms of the transport of ozone and its precursors, southerly winds will advect air into the region from marine or predominantly rural areas. The background levels of ozone and its precursors in this situation will be considerably different from air masses advected from the industrial areas to the west or northwest even though the temperature conditions may be similar. Recent aircraft observations have shown that a deep layer of ozone enriched air can be observed in cases of northwesterly transport (Morales *et al.*, 1994).

In terms of likely forecast skill, the regional CART analysis shows reasonable skill forecasting null ozone events and distinguishing the most severe events. Its resolution is limited in cases of "warm" temperature with a tendency to underpredict. The tendency of the CART forecasts to underpredict ozone will likely result in fewer "false alarms". The strength of temperature as a predictor suggests that a regression approach may be useful for ozone forecasting. The difficulties with the CART approach in the mid-range cases suggests that an expert modification of both CART and regression analysis may also be helpful. In particular, the subjective analysis can provide information on the position of surface and upper air ridges and transport effects associated with them.

3. THE BALTIMORE FORECASTING PROGRAM

The goal of the Baltimore forecast program is to provide a simple forecast model for ozone that can be used on a daily basis at the Maryland Department of the Environment (MDE). The forecast will be used to issue health advisories and drive episodic emissions reduction strategies. The former will require a short-term, fairly precise, forecast while the latter will require an additional longer range forecast to accommodate those reductions that cannot be put in place in less than 24 h. Because the long-range forecast is geared to industrial and governmental use, this forecast uses the NAAQS standard as a categorical forecast level.

The Baltimore ozone forecasts were prepared daily at the University of Maryland between 1200 and 1600 UTC for the period June 15–July 31 and provided a peak 24 h $[O_3]$ prediction (in ppbv) as well as a 2–3 day outlook for $[O_3]$ in excess of 120 ppbv. Because the forecasts are issued before the 1200 UTC NWS meteorological forecast model runs are completed and disseminated, the 0000 UTC model runs are used for forecast input. These meteorological forecasts thus verify roughly 36 h after issuance—late afternoon of the following day. For surface variables, the Model Output Statistics (MOS) based on the Nested Grid Model (NGM) are used. For upper air conditions, the NGM or Medium Range Forecast (MRF) are used depending on their availability at the forecast site. The upper air predictors (850 mb) are determined by a gridded area average from the appropriate model level using the GRads display package (Doty, 1992). For several periods early in the forecast program, the acquisition of NGM and MRF data is interrupted. For these days, the 1200 UTC sounding data from IAD is coupled with expected advection conditions to give a forecast of 850 mb conditions. The data from the meteorological models are input to the two ozone forecast models and the results used for preparation of the subjective forecast.

Table 4. Parameters and basic results for the ozone regression analysis for the Baltimore metropolitan area

Parameters	Minimum tolerance	0.100
	α to enter	0.100
	α to remove	0.100
	N	731
Results	Multiple R	0.765
	Squared multiple R	0.585
	STD error	19.3

For the summer 1993 program, three ozone forecast models are used for the 24 h forecast. First, a localized CART forecast, second, a simple regression-based forecast and, finally, a subjective, or expert, forecast. The subjective analysis alone is used to develop the 2–3 day outlook. The Baltimore regression analysis utilizes ozone observations from 1983 to 1990 (June–August) at four locations in the Baltimore area (Essex, Padonia, Fort Meade and Garrison, Fig. 2). A forward stepwise regression is performed with parameters set as noted in Table 4. The meteorological data for the regression analysis are similar to that utilized in the regional CART analysis (Part II) with the exception of only limited upper air data (850 mb only). The regression analysis resulted in a set of six predictors. The weight and average values of each are given in Table 5. An additional run utilizes upper air data from 950 to 900 mb. The differences between the two runs are slight and because data for 850 mb is easier to acquire routinely, the simpler regression model is used for the forecast. An additional run with ozone data from the Washington metropolitan area selected the same main predictors with the exceptions that 850 mb geopotential height (Z850) is replaced by 850 mb wind speed (WS850) and dew point temperature is added.

The local CART analysis utilizes the same data set as the regression approach and is also dominated by the 1800 UTC temperature ($T18$) and sky cover (TSKC18). Additional variables of importance are 850 mb temperature ($T850$) and WS850 as well as 1000 UTC temperature ($T10$) and pressure ($P10$). The CART analysis resulted in nine terminal nodes (Table 6). As in the regional analysis, the strongest and weakest ozone cases are fairly well defined. Of the cases grouped in TN8 ($T18 > 92.5°F$ and $WS850 < 6.5\ m\ s^{-1}$) 85% contain ozone violations

Table 5. Summary of predictors used in the ozone regression analysis for the Baltimore metropolitan area

Predictor	COEF	STD error	STD COEF	TOL	Mean	ΔO_3 for Δ 5% predictor
K	−136.1	33.5	0.0			
$T10$	−1.88	0.20	−0.4	0.34	67.1	6.3
$T18$	−2.53	0.18	0.61	0.29	83.0	10.5
TSKC18	−1.07	0.27	−0.12	0.56	5.0	0.3
WS18	−1.68	0.23	−0.18	0.90	8.9	0.7
$Z850$	0.08	0.02	0.10	0.80	1536	6.1
$T850$	2.25	0.41	0.25	0.26	14.7	1.7

Surface variables are given by type (T = temperature, TSKC = sky cover, WS = wind speed) and time of observation (18 = 1800 UTC), upper air variables by type (Z = geopotential height) and level (850 = 850 mb). Regression coefficients (COEF), standard error (STD error), standard coefficients (STD COEF) and tolerance (TOL) are given for each predictor along with a mean value. The measure ΔO_3 refers to the change in predicted $[O_3]$ induced by a 5% variation in predictor value from the mean.

Table 6. CART results for the Baltimore metropolitan area; ozone concentrations are peak daily values and meteorological variables are defined as in previous tables

Terminal node	N	$[O_3] >$ 120 ppbv	Median $[O_3]$ (ppbv)	BWI mean T (°F, 1800 UTC)	BWI mean wind speed (kts, 1800 UTC)	IAD mean 850 mb wind speed (m s^{-1}, 1200 UTC)
1	155	1	59	72.8	8.9	7.8
2	176	5	73	81.8	9.0	8.5
3	119	18	95	81.8	7.8	4.6
4	9	0	57	81.0	9.3	8.2
5	33	21	128	88.2	5.5	5.7
6	62	14	104	88.2	10.7	7.0
7	116	19	92	88.9	9.5	8.1
8	39	33	134	94.9	8.3	4.1
9	27	6	107	95.3	11.3	11.0
Total	736	117				

while those in TN6 (85.5°F < $T18$ < 92.5°F, T_d < 69.5°F and weak winds) have a 65% occurrence rate. Conversely, those in TN1 and TN2 ($P10$ < 1010 mb and $T18$ < 85.5°F) have a less than 2% rate of occurrence. However, the two strongest nodes contain only 46% of all strong ozone occurrences. The remaining strong ozone days are sprinkled throughout the "warm" nodes (TN3, TN6, and TN7). It is of interest that the warmest node (TN9) contains a very low rate of severe ozone cases due to strong winds. In general, the results of the local CART analysis are similar in result to the regional CART with a problem of underprediction in the less severe ozone terminal nodes.

The forecast meteorological variables used by both the CART and the stepwise regression analyses vary in terms of type and expected skill and can constrain the ozone forecast. In the last column in Table 5, the effect on the regression O_3 forecast of a 5% variation in predictor values is given. The surface variables are derived from the NGM-MOS (Carter *et al.*, 1989). The MOS forecasts are specific for BWI and expected to be close to observations for the 24–36 h time period. However, several parameters are expressed in different form in the MOS than in the data base for the regression equation and may be a source of confusion. For example, the MOS sky cover forecasts are categorical (clear, scattered, broken, overcast) and do not correspond to the observed sky cover report expressed in tenth's. However, the slight weight to sky cover means that even a variation of 30% will result in only a 3–4 ppbv difference in forecast $[O_3]$. Similarly, the wind speed forecast given by the MOS is a mean value for the 6 h forecast period while the observation period for NWS data is on the order of 1–2 min. The spectra of wind variance has been found to have a good deal of energy in the 2–3 min period so that local observations may frequently vary considerably from six hour averages (Panofsky and Brier, 1965; Powell, 1993).

The upper air parameters are derived directly from the NGM and MRF models. In terms of the regression analysis, the 850 mb height variable ($Z850$) is probably the biggest expected source of error. The natural variability of geopotential height is quite large with heights ranging from 1500–1600 m in this season. In addition, the observations are made by single flight radiosondes which introduces an instrument error problem. With this in mind, an error of only 10–20 m will result in ozone variations of 8–16 ppbv.

The subjective forecast is based on a modification of the CART and regression analyses. Several criteria that can affect $[O_3]$ but are not explicitly contained in the objective forecasts are used in the preparation of the subjective forecasts. First, synoptic scale pressure and geopotential height fields are analyzed to determine the position of surface and upper air ridges. The positioning of a surface anticyclone to the south and west of Baltimore with 850 mb winds from the west or northwest is used to adjust ozone forecasts above 120 ppbv. Second, the likelihood of strong and widespread convection prior to the time of maximum ozone concentration (typically 2000–2200 UTC) is used to adjust forecast $[O_3]$ downward. Although the frequency of high ozone concentrations are not strongly influenced by day of the week, the most severe events tend not to occur on Sundays. Therefore, Sunday forecasts were modified slightly downward in high ozone conditions. As noted above, frontal passages and abrupt wind shifts may affect ozone levels although the direction of the effect is difficult to determine. In cases where frontal zones approached from the north (the "back door" front common to this region), the ozone forecast was adjusted downward. This is based on the expectation of cool temperatures and low concentrations of O_3-precursors associated with the advection of maritime air. "Back door" frontal zones are frequently associated with a ridge of high pressure to the west of Baltimore. The latter situation is, as noted in the subjective analysis, associated with high ozone episodes. Therefore, the gradient of ozone across this type of front is strong and, as found in several cases in June, 1993, the failure of the front to cross Baltimore resulted in large subjective forecast errors. Finally, if significant precipitation or strong winds are forecast to occur with the passage of a front, a downward adjustment is made.

4. FORECAST RESULTS

The 24 h ozone forecasts were verified against peak $[O_3]$ reported at all monitors in the Baltimore area. These included several sites initiated after the climatological data base used here (1983–1990). These include three stations well outside the Baltimore suburbs (Fair Hill, Aldino and Millington) that may be influenced by Philadelphia emissions. To gauge this effect, the forecasts were also verified against the four local stations used in the preparation of the regression and CART forecasts (Garrison was discontinued prior to 1993). There were only slight differences in forecast skill reported between the two groups and the results from the larger group are reported here.

Basic statistical measures for the forecasts are shown in Table 7. The regression model correlates best although, as may be expected, it does not fair well at the extreme values. The CART approach does not correlate well. The subjective approach is moderately well correlated but suffers from occasional large lapses driven by missed forecasts of frontal positions and timing of convective activity. For $[O_3]$ in excess of 100 ppbv, the standard error for all forecasts is reduced.

Because the forecast program is geared to the threshold of the NAAQS, an alternative method of determining forecast success is by investigating categorical results, that is, the success of forecasts of extreme ozone conditions. The results are based on the contingency table given in Table 8. The simplest measure is Probability of Detection (POD) which measures the ability of the model to accurately forecast an event (see, Appendix A for more details on skill measures). Other useful measures are those that measure overprediction (False Alarm Rate (FAR)), underprediction (Miss Rate (MISS)), and correct null, or low ozone forecasts (CNULL). Three common skill score measures that combine the simpler measures are also given (CSI, TSS, S). For forecasts of 120 ppbv, no method shows great skill although the subjective forecast rises to a 50% POD (Table 9). Because all approaches tend to underpredict $[O_3]$, better results are obtained for forecasts of $[O_3]$ in the moderate to high (110 ppbv or greater) range, particularly when verified against higher observed ozone. The best predictions of ozone violations are obtained by a subjective forecast of $[O_3] > 115$ ppbv verifying against observed $[O_3]$ in excess of 120 ppbv. This measure achieved a 75% POD of an ozone exceedance.

The sources of forecast error include both the limitations of the forecast models as well as errors in the input data on which the forecasts are based. The errors in the forecasted meteorological data are given in Table 10. In general, the forecasts are in the same range as observations although Z850 is consistently underpredicted. For warm periods ($T18 > 90°F$) temperature is underpredicted with a mean difference of 3.6°F leading to an $[O_3]$ underprediction of 9 ppbv. The effect of meteorological forecast data errors on the categorical forecast are seen when the ozone forecast models are initialized after the fact with observations as the predictors (perfect forecast) (Table 11). Using observed meteorological data improves all measures with the CART model, which is heavily reliant on the initial temperature split, showing the greatest gain.

One limitation of the forecast model is the failure to explicitly include stability parameters such as the depth or strength of the morning inversion layers. The persistence of the morning inversion may play a role in increasing $[O_3]$. Because these inversions are highly variable, and may depend on mechanical effects such as large scale subsidence, there is no routinely available forecast product that can adequately assess it in the 36–48 h time frame. Using sounding data from the 1993 cases, we can backcast to see if there is a strong effect associated with inversion height or strength. For the summer 1993 cases, a regression analysis was performed using the predictors from the localized regression with a set of stability parameters

Table 7. Basic statistics for Baltimore ozone forecasts

	All days ($N = 46$)			$[O_3] > 100$ ppbv ($N = 24$)		
Type of forecast	Mean	CORR	Error	Mean	CORR	Error
CART	94.9	0.44	23.0	98.4	0.53	13.8
REGR	97.2	0.73	17.5	105.4	0.56	13.4
SUBJ	97.1	0.67	19.1	108.8	0.47	14.3
CART-PF	100.6	0.58	17.9	108.5	0.49	14.2
REGR-PF	106.1	0.79	20.9	115.7	0.58	13.3
OBS	102.2			121.2		

REGR refers to forecasts based on the ozone regression equation. SUBJ refers to forecasts made by subjective, or expert analysis and CART refers to forecasts made using the CART analysis. Those forecasts suffixed PF (Perfect Forecast) refer to ozone forecasts which were prepared after the fact using observed, rather than forecast, meteorological parameters.

Table 8. Contingency table used for verification of ozone forecasts

	Forecast yes	Forecast no
Observed yes	A	B
Observed no	C	D

added. These included the pressure level of the surface-based inversion at 1200 UTC, the differences in temperature and potential temperature (θ) between the surface and the inversion level as well as between the surface and the NWS mandatory levels of 925 and 850 mb. Finally, the strength of the inversion, as denoted by the product of the temperature difference and height of the inversion was calculated. The control run, using only the previous predictors, produced a regression equation that included $T18$, WS18 and $Z850$, a correlation coefficient of 0.803 with a standard error of 15.7 ppbv and explained variance of 64%. The inclusion of the stability data added the strength of the inversion as an additional predictor and increased the correlation coefficient to 0.860 with a standard error of 14.0 ppbv and explained variance of 74%. Because of the difficulty in forecasting the

Table 9. Forecast skill based on categorical measures

		Measures of forecast skill						
Type of forecast	Threshold	POD	FAR	MISS	CNULL	CSI	TSS	S
REGR	120	0.17	0.00	0.83	1.00	0.17	0.17	0.23
CART	120	0.30	0.00	0.70	1.00	0.30	0.30	0.40
SUBJ	120	0.50	0.33	0.50	0.91	0.40	0.41	0.45
SUBJ	115	0.64	0.25	0.36	0.91	0.53	0.55	0.54
REGR	110	0.47	0.27	0.53	0.90	0.40	0.37	0.40
CART	110	0.60	0.40	0.40	0.89	0.43	0.49	0.47
SUBJ	115	0.75	0.25	0.25	0.91	0.60	0.66	0.62
OBS	120							

Measures of forecast skill are defined in the text and Appendix. As in Table 7, ozone forecasts include regression (REGR), CART (CART) and subjective (SUBJ). Forecasts are verified against observations (OBS) for a given threshold (e.g. REGR > 120 verifies a regression forecast of greater than 120 ppbv against observations of greater than 120 ppbv). Cases in which ozone forecasts are verified with observations at a different threshold are noted.

Table 10. Accuracy for forecast meteorological parameters used as input to the ozone forecast models

Parameter	Forecast mean	Observed mean	Standard error	CORR
Surface temperature (1000 Z)	69.2	69.3	3.4	0.76
Surface temperature (1800 Z)	86.0	87.3	3.2	0.86
Sky cover (1800 Z)	3.1	4.6	2.9	0.40
Wind speed (1800 Z)	8.6	8.3	2.6	0.36
850 mb temperature (1200 Z)	16.1	15.9	1.8	0.70
850 mb geopotential height (1200 Z)	1539.0	1552.0	28.0	0.67

Table 11. Forecast skill for O_3 forecasts utilizing observed meteorological data (perfect forecast); thresholds are used as in Table 9

		Measures of forecast skill						
Type of forecast	Threshold	POD	FAR	MISS	CNULL	CSI	TSS	S
REGR	120	0.50	0.40	0.50	0.88	0.38	0.38	0.40
CART	120	0.50	0.14	0.50	0.97	0.46	0.47	0.54
REGR	110	0.65	0.31	0.35	0.83	0.50	0.47	0.43
CART	110	0.59	0.33	0.41	0.83	0.45	0.42	0.40
REGR or SUBJ,	115	0.83	0.38	0.17	0.75	0.56	0.58	0.38
OBS	120							
REGR and SUBJ,	115	0.58	0.22	0.42	0.91	0.50	0.49	0.51
OBS	120							
REGR,	115	0.75	0.36	0.25	0.85	0.53	0.60	0.50
OBS	120							

Table 12. Forecast skill based on categorical measures for the two and three-day O_3 forecasts

Type of forecast	Measures of forecast skill						
	POD	FAR	MISS	CNULL	CSI	TSS	S
2-day FCST	0.33	0.33	0.67	0.94	0.29	0.27	0.35
3-day FCST	0.36	0.33	0.64	0.93	0.31	0.29	0.36

specific level and strength of the following day's inversion, an additional regression run was made using only stability data from 925 and 850 mb. In this case, the inversion parameter is replaced by 850 mb data with similar results including a correlation coefficient of 0.865 with a standard error of 13.9 ppbv and explained variance of 75%. When this parameter is tested with the original 1983–1990 database, however, it was not to be a strong predictor.

The long-range forecast results are not good (Table 12). The POD is fairly small (0.33–0.36) with a corresponding high miss rate (0.64–0.67) and relatively low skill scores. However, the ability of the MRF to forecast longer term (0–10 days) heat waves made it possible to forecast a developing multi-day ozone event (July 3–9) accurately four days in advance. Daily forecasts of $[O_3]$ in excess of 120 ppbv within the heat-wave period are less successful and the ability to predict single day events was also spotty. It appears that the current skill of the multi-day forecast is the ability to predict multi-day ozone episodes in the context of extended heat waves.

5. SUMMARY AND CONCLUSIONS

An effort is undertaken to predict ozone episodes in the Baltimore region using several approaches based on meteorological parameters. The use of a CART analysis for the B/W region, using historical data, shows skill at distinguishing strong and weak ozone cases but is unable to clearly separate the majority of ozone events into separate clusters. In operational use in the Baltimore area, CART did not well predict ozone events. The CART analysis is characterized by poor correlations with observations and high standard error (23 ppbv) along with a high miss rate (0.70) and low skill scores (0.30–0.40). The sources of error in the CART analysis include limitations of the database on which it was initialized as well as poor forecasts of mid-day temperatures. The CART analysis is strongly dependent on accurate temperature forecasts and forecast skill improved by 50% with the use of accurate temperature data.

The regression analysis performed better than the CART model with a fairly high correlation coefficient (0.73) and a standard error of 17.5 ppbv. The regression model consistently underpredicts $[O_3]$ with a low bias of 12 ppbv. This is due in most part to poor temperature forecasts. With accurate temperature data, the skill scores for the regression model more than doubled. The poor temperature forecasts are most damaging during warm periods with a low bias of 4°F for days with temperatures greater than 90°F. A slight low bias still remains even with perfect meteorological data and the best success is obtained when regression forecasts of greater than 115 ppbv are verified against observations of greater than 120 ppbv. In this case the POD rises to 0.75 with skill scores in the range 0.50–0.60.

The subjective analysis results for all forecasts track the regression approach in terms of correlation and standard error values. However, for the stronger ozone episodes, the subjective forecast performs best of all approaches with skill scores of 0.40–0.45. The subjective approach also suffers from a low bias and the best results are obtained with subjective forecasts of greater than 115 ppbv verified against observations of greater than 120 ppbv. In that case, the skill scores rise to the range 0.60–0.66 with a POD of 0.75.

There are a number of areas in which these models can be improved for future forecasts. First, the use of local NWS temperature forecasts for warm periods will remove some degree of underprediction. During the warm periods of 1993, the NGM MOS was consistently modified upward by local NWS forecasters with better results. Second, the regression and CART models will be initialized with data from 1987 to 1993. This may introduce improvements related to changes in the emissions base during the recent past. In addition, more ozone monitor data are available for the Baltimore area for this period. Finally, the subjective forecast will be modified during warm periods to account explicitly for the low bias. In the longer term, the most improvements will be made when a strong set of O_3-precursor observations can be added to the climatological database.

REFERENCES

Breiman L., Friedman J. H., Olshen R. A. and Stone C. J. (1984) *Classification and Regression Trees.* Wadsworth, Belmont, CA.

Burrows W. R. (1991) Objective guidance for 0–24-hour and 24–48-hour mesoscale forecasts of lake-effect snow using CART. *Wea. Forecasting* **6**, 357–378.

California Statistical Software (1991) *CART.* Lafayette, California.

Carter G. M., Dallavalle J. P. and Glahn H. R. (1989) Performance of the national meteorological center's numerical weather prediction system. *Wea. Forecasting* **4**, 401–412.

Doswell C. A. III, Jones R. and Keller D. L. (1990) On summary measures of forecast skill in rare event forecasting based on contingency tables. *Wea. Forecasting* **5,** 576–585.

Doty B. E. (1992) *Using the Grid Analysis and Display System (GrADS)*, Center for Ocean–Land–Atmosphere Interaction, University of Maryland, College Park.

EPA (1991a) National Air Quality Emissions Trends Report, EPA-450/4–91-003, U.S. Environmental Protection Agency, Office of Air Quality Policy and Standards, Research Triangle Park, North Carolina.

EPA (1991b) *Regional Ozone Modeling for Northeast Transport,* EPA-450/4-91-002a, U.S. Environmental Protection Agency, Office of Air Quality Policy and Standards, Research Triangle Park, North Carolina.

EPA (1991c) Guidelines for Regulatory Application of the Urban Airshed Model, EPA-450/4-91-013, U.S. Environmental Protection Agency, Office of Air Quality Policy and Standards, Research Triangle Park, North Corolina.

Environmental Science Services Administration (ESSA) (1968) *Climatic Atlas of the United States,* U.S. Dept. of Commerce, Washington, D.C.

Horie Y. (1987) *Ozone Episode Representativeness Study for the South Coast Air Basin,* Valley Research Corp., prepared for the South Coast Air Quality Management District, El Monte, California.

Lee R. L. and Passner J. E. (1993) The development and verification of TIPS: an expert system to forecast thunderstorm occurrence. *Wea. Forecasting* **8,** 271–280.

Morales R., Ryan W. F., Doddridge B. G. and Dickerson R. R. (1994) Air quality measurements in the Baltimore metropolitan area during the summer of 1993. *EOS, Trans. Am. Geophys. Union* **75,** 88.

Moy L. A., Ryan W. F., Doddridge B. G. and Dickerson R. R. (1994) Relationship between back trajectories and tropospheric trace gas concentrations in rural Virginia. *Atmospheric Environment* **28,** 2789–2800.

National Research Council (1991) *Rethinking the Ozone Problem in Urban and Regional Air Pollution.* National Academy Press, Washington, D.C.

NAVAIR (1966) *Components of the 1000 mb Winds of the Northern Hemisphere,* NAVAIR-50-1C-51. U.S. Government Printing Office, Washington, D.C.

Pagnotti V. (1987) A meso-meteorological feature associated with high ozone concentrations in the Northeastern United States, *JAPC A.* **37,** 720–722.

Panofsky H. A. and Brier G. W. (1965) *Some Applications of Statistics to Meteorology.* College of Mineral Industries, The Pennsylvania State University, State College, Pennsylvania.

Pierrehumbert R. T. (1986) Lee cyclogenesis. In *Mesoscale Meteorology and Forecasting* (edited by Ray P. S.), pp. 493–515. American Meteorological Society, Boston.

Powell M. D. (1993) Wind measurement and archival under the Automated Surface Observing System (ASOS): user concerns and opportunity for improvement. *Bull. Am. Meteor. Soc.* **74,** 615–629.

Schulze R. H. (1993) The 20-year history of the evolution of air pollution control regulation in the U.S.A. *Atmospheric Environment* **27B,** 15–22.

Seinfeld J. H. (1988) Ozone air quality models: a critical review. *JAPCA.* **38,** 616–645.

Seinfeld J. H. (1989) Urban air pollution: state of the science. *Science* **243,** 745–752.

Trenberth K. E., Branstator G. W. and Arkin P. A. (1988) Origins of the North American drought of 1988. *Science* **242,** 1640–1645.

Vukovich F. M. and Fishman J. (1986) The climatology of summertime O_3 and SO_2 (1977–1981). *Atmospheric Environment* **20,** 2423–2433 (1986).

Vukovich F. M., Bach Jr W. D., Crissman B. W. and King W. J. (1977) On the relationship between high ozone in the rural surface layer and high pressure systems. *Atmospheric Environment* **11,** 967–983.

APPENDIX: VERIFICATION METHODOLOGY

The verification of the ozone forecasts utilized a standard contingency table of the type displayed in Table 8. Forecast skill can be expressed in a number of ways. The Probability of Detection (POD) measures the percentage of ozone events that were correctly forecast and is given by:

$$\mathrm{POD} = \frac{A}{A+B}.$$

The Miss Rate (MISS) measures the rate at which ozone events occurred but were failed to be forecast and is given by

$$\mathrm{MISS} = 1 - \mathrm{POD} = \frac{B}{A+B}.$$

The False Alarm Rate (FAR) measures the tendency of the zone forecast to overpredict ozone occurrences and is given by

$$\mathrm{FAR} = \frac{C}{C+A}.$$

In addition to forecasts of ozone occurrences, an alternative measure is the skill at which non-events are forecast. This is a measure of the forecast skill at predicting "clean" days and is given by the Correct Null Forecast (CNULL):

$$\mathrm{CNULL} = \frac{D}{D+C}.$$

There are a number of more detailed skill scores that are useful. The Critical Success Index (CSI or Threat Score) combines forecast occurrences and observed occurrences without regard to successful null forecasts. This is given by

$$\mathrm{CSI} = \frac{A}{A+B+C}.$$

The True Skill Score (TSS or Hanssen-Kuipers Skill Score) includes the success of null forecast in the form of a ratio of observed skill to perfect forecast skill. This measure is not dependent on the relative frequency of occurrence and non-occurrence or the number of trials and can be expressed as

$$\mathrm{TSS} = \frac{(AD)-(BC)}{(A+B)(C+D)} = \mathrm{POD} + \mathrm{CNULL} - 1.$$

If all forecasts are correct, TSS = 1, if all forecasts are incorrect, TSS = − 1 (Lee and Passner, 1993).

For events that are rare, the Heidke Skill Score (S) is often used. This score is a measure of the skill of a set of forecasts compared to the skill of a random forecast (Doswell *et al.*, 1990).

$$S = \frac{2(AD-BC)}{B^2 + C^2 + 2AD + (B+C)(A+D)}.$$