4.15 AN EVALUATION OF OZONE FORECAST METHODS FOR THE EASTERN UNITED STATES: IS THERE ROOM FOR IMPROVEMENT?

Scott A. Jackson*[1], Alan C. Chan[2], Craig E. Lengyel[2], Richard A. Wayland[1], John E. White[1], and Timothy S. Dye[2]

[1] U.S. EPA/Office of Air Quality Planning and Standards
[2] Sonoma Technology, Inc.

1. INTRODUCTION

Air quality forecasting has become more common throughout the United States over the past 10 years (Dye et al., 2004). What began with a few states reporting current ozone conditions evolved into a nationwide network of state and local agencies producing forecasts for over 300 cities. More specifically, ozone forecasts are produced from the beginning of May through the end of September, a period typically considered the "ozone season" in the United States.

Most agencies began forecasting ozone as the only pollutant. The promulgation of new 24-hr $PM_{2.5}$ standards in 1997 (Federal Register, 1997), in addition to numerous recent studies showing increasing harm caused by $PM_{2.5}$ to human health (Peters et al., 2001), has recently led many agencies to begin forecasting particle pollution in addition to ozone. As multi-pollutant prediction gained momentum, a new idea of "air quality" rather than "ozone" forecasting began. While the scope of this paper is to examine the performance of ozone prediction tools and methods, it is important to note that air quality forecasting (which includes multiple pollutants) is becoming more prevalent (Wayland et al., 2004).

Many agencies are forecasting air quality at various levels of government. A byproduct of the decentralized nature of this forecasting community is a variety of methods for forecasting air quality and, more specifically, ozone. The complexity and effectiveness of the tools available depend greatly on the maturity of a particular forecasting program. Forecasting methods range from simple rules of thumb to statistical methods to more complex real-time air quality models. Wayland et al. (2002) noted that uniformly assessing the accuracy of the wide range of tools and methods is a challenge. In addition, because nationwide air quality forecasting is only a recent development, the current state of air quality forecasting has been compared to the state of weather forecasting 20 years ago.

This paper evaluates the performance of several ozone forecasting methods (statistical models, real-time photochemical models, and human forecasting). In addition to performance assessment, this paper recommends forecast performance goals for air quality forecasting programs across the country.

Corresponding author address: Scott A. Jackson, U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711; email: jackson.scott@epa.gov

1.1 *Distribution of Air Quality Forecasts*

State and local agencies are the authority for issuing air quality forecasts and warnings for protecting public health. Air quality forecasts are available regionally through web sites and other outreach sources and nationwide through the U.S. Environmental Protection Agency's (EPA) AIRNow program (www.airnow.gov). AIRNow is a voluntary partnership between EPA and regional, state, and local agencies that began in 1997 with a goal to provide current air quality conditions and forecasts for the United States.

Forecasts are submitted to AIRNow's Data Management Center (DMC) and then distributed to numerous organizations. One major destination of forecast data is Weather Service Providers (WSPs) who provide data to television and print media outlets. The media are an effective way to disseminate forecast and related health information to the public. While AIRNow plays a significant role in distributing both real-time and forecast air quality information, many regional, state, and local programs have a well-established infrastructure for transmitting their data to local media and the public.

1.2 *Selection of Forecast Cities*

Six cities in the eastern United States were selected for the purpose of evaluating agency air quality forecasting performance. The basic criteria for selection were that (1) each city must be within the domain of the real-time air quality model and (2) the agency responsible for a city's air quality forecasting has several years of experience forecasting ozone. While not required, preference was given to agencies using a statistical tool to aid human forecasts. The cities were also chosen based on geographic diversity to capture variations in weather patterns. Figure 1 shows the cities that met the criteria for selection: Atlanta, Georgia; Baltimore, Maryland; Charlotte, North Carolina; Columbus, Ohio; Boston, Massachusetts; and Richmond, Virginia.

2. TYPES OF FORECASTING METHODS

Three types of forecasting methods were compared as a measure of performance: human forecasting, statistical methods, and real-time air quality modeling. While human forecasting is the most common method, statistical and numerical models have become useful and respectable guidance tools to support human air quality forecasts.

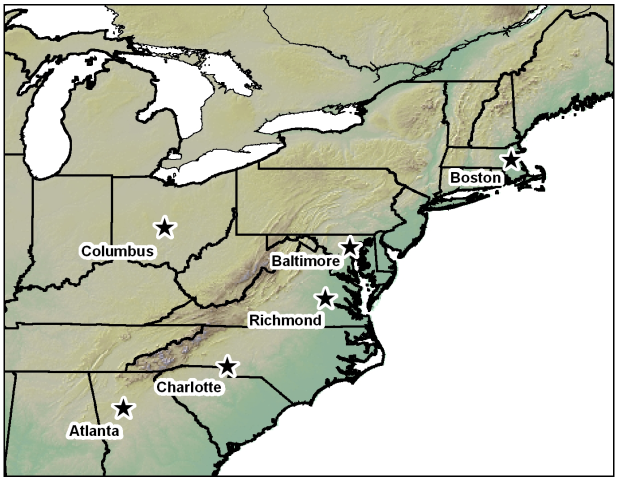

Figure 1. Six cities within the NOAA model domain included in this study.

2.1 *Human Forecasting*

There is no explicit definition of human air quality forecasting. The most common method used by human forecasters in the six cities that were evaluated is the phenomenological/intuition method. This method of forecasting relies heavily on analyzing and conceptually processing air quality and meteorological data to formulate an ozone prediction (U.S. Environmental Protection Agency, 2003). Forecaster experience plays a large role in the phenomenological method.

For forecasting method comparison in this study, it is assumed that the human forecasters in each city have equal access to the tools to which human forecasting is being compared in addition to their own forecasting experience. Since forecaster experience and training vary, directly comparing human forecasting skill across the cities is difficult.

2.2 *Statistical Methods*

The statistical methods evaluated for the select cities consist of multiple linear regression equations. A regression equation uses relationships between specific variables to predict outcomes (U.S. Environmental Protection Agency, 2003; Comrie, 1997). In the case of air quality, specific meteorological variables are identified that correlate with ozone concentrations. Most regression equations use wind speed, maximum surface temperature, aloft temperatures, and cloud cover as predictor variables.

2.3 *Real-Time Air Quality Models*

Air quality, grid-based models have, for many years, been available for regulatory purposes to evaluate emission control scenarios. Only in recent years have universities, national centers, and private sector organizations begun running real-time air quality models to support operational ozone forecasting. The two modeling efforts discussed in this paper are from the Meteorological Service of Canada (MSC) and the National Oceanic and Atmospheric Administration (NOAA).

In summer 1999, the MSC began a pilot ozone forecasting program for eastern Canada only, using the Canadian Hemispheric and Regional Ozone NO_x System (CHRONOS). The CHRONOS model domain now covers most of North America including all the contiguous United States. The model has a 21-km grid resolution and runs twice a day at (0Z and 12Z). Additional details about the model can be found in Venkatesh (2003).

In the United States, NOAA and EPA developed an experimental real-time air quality forecast model. The model uses NOAA's National Weather Service (NWS) operational model, Eta, for its meteorological input. The air quality component of the modeling system is the EPA Community Multi-Scale Air Quality (CMAQ) model (McQueen et al., 2004). The model was run as a prototype for the northeastern United States domain during summer 2003 (Fig. 1) with a horizontal grid spacing of 12 km. Started in early May 2004, the modeling system became operational by late summer.

3. ASSESSMENT AND PERFORMANCE

Weather conditions during summer 2004 were generally cooler and cloudier than normal in the eastern United States. These conditions produced a below average number of high-ozone episodes. Nevertheless, evaluating forecast performance is important to establish a baseline for future forecasting efforts and to understand forecast performance in a relatively clean ozone year.

We evaluated next-day forecasts from May 1 to September 30, 2004, for the Air Quality Index (AQI) categories (see www.airnow.gov/aqi.html for category descriptions). Two thresholds on the AQI scale were selected: Moderate and above (≥65 ppb 8-hr average concentration) and Unhealthy for Sensitive Groups (USG) and above (≥85 ppb 8-hr average concentration). The second threshold was chosen because it is the threshold for the national ozone standard and corresponds to action days (e.g., ozone alerts, code-red days, Spare The Air days, etc.) typically issued by air quality agencies.

Forecasts were evaluated as to whether the predictions were above or below the thresholds. The following verification statistics were used:

- **Percent correct (PC) –** Shows the percent of forecasts that correctly predicted the event or non-event.

- **False alarm rate (FAR) –** Shows the percent of time a forecast of high pollution (at or above the threshold) was wrong.
- **Probability of detection (POD) –** Indicates the ability to predict actual high-pollution events (i.e., at or above the threshold).
- **Critical success index (CSI) –** Shows how well high-pollution events were predicted; it is unaffected by a large number of correctly forecasted, low-pollution events.
- **Bias –** Indicates, on average, whether forecasts are under-predicted (false negatives) or over-predicted (false positives).

Forecast statistics were calculated for the six cities at the two AQI thresholds—Moderate and above and USG and above as shown in Tables 1 and 2. Each forecasting method—human experience, statistical tools, and the NOAA and CHRONOS models—was then objectively compared to the others. Results from the CHRONOS model will be shown at the AMS conference.

In general, statistics show that human forecasters performed better at both thresholds (Table 3). Specifically, human forecasts, on average, were the least biased and had the lowest FAR. These forecasts also correctly predicted 96% and 80% of all events at the first and second thresholds, respectively. When only high-ozone events were considered, human forecasters were able to predict almost 80% (POD) of the Moderate AQI and above events as shown in Table 1. At the USG and above threshold (Table 2), about one-third of the high days were detected by human forecasters.

The NOAA model correctly predicted 47% of the USG and above days. But it over-predicted more frequently than both statistical and human forecasts at the same threshold, yielding a bias of 3 and a FAR over 80%. This is especially important because many forecasters are beginning to depend on NOAA model guidance to aid their forecasts. In addition, regional, state, and local agencies across the country spend significant resources on outreach efforts during forecasted poor air quality days. Regionally, the NOAA model was able to predict high-ozone days (USG and above) better in Atlanta, Charlotte, and Richmond than in other cities used in this study.

Statistical tools, which were not available in all cities, were generally outperformed by human forecasters. However, the statistical tool for Baltimore correctly forecasted 80% of the high-ozone days during 2004 and performed better than the human forecaster.

4. SUMMARY AND RECOMMENDATIONS

Results of this study showed that human forecasts were generally more accurate in predicting high-ozone days based on results from the six cities for summer 2004. At the same time, photochemical models and statistical tools were able to detect high-ozone days in certain areas in this study and can provide valuable guidance to air quality forecasters.

Air quality forecasting, compared to weather forecasting, is still in its infancy, and the air quality forecasting community faces several challenges in achieving more accurate forecasts: (1) how to determine which forecast tools work best in certain situations; (2) how to create a consensus forecast by using various guidance prediction; (3) how to forecast for multiple pollutants; and (4) how to evaluate and learn from missed forecasts.

Ozone forecast performance can be improved. We surveyed the air quality forecasting and pubic outreach communities, and representatives of those entities did not have rigid expectations in terms of forecast performance. We recommend the following performance goals for forecasting USG and above ozone days: a FAR less than 25% and a POD of greater than 80%. We also suggest performing the analysis again for a high-ozone season and for other pollutants—for example fine particles ($PM_{2.5}$)—which will better reflect each forecast method's ability to detect high-pollution days.

The comparisons provided in this paper serve as a baseline for forecast performance. As more air quality forecast tools are developed and photochemical models are revised and improved, they will continue to aid air quality forecasters to achieve these goals.

5. REFERENCES

Comrie A.C., 1997: Comparing neural networks and regression models for ozone forecasting. *J. Air & Waste Manag. Assoc.* **47**, 653-663.

Dye T.S., A.C. Chan, C.B. Anderson, D.E. Strohm, R.A. Wayland, and J.E. White, 2004: From raw air quality data to the nightly news: an overview of how EPA's AIRNow program operates. Paper presented at the *2004 American Meteorological Society Annual Meeting, AMS Sixth Conference on Atmospheric Chemistry: Air Quality in Megacities, Seattle, WA, January 11-15,* by Sonoma Technology, Inc., Petaluma, CA, and U.S. Environmental Protection Agency, Research Triangle Park, NC (STI-902107-2405).

Federal Register, 1997: Review of national ambient air quality standards for particulate matter, 40 CFR Part 50, July 18. Available on the Internet at http://www.epa.gov/EPA-AIR/1997/October/Day-23/a28143.htm.

McQueen J., P. Lee, M. Tsidulko, G. DiMego, T. Otte, J. Pleim, J. Young, G. Pouliot, B. Eder, D. Kang, K. Schere, J. Gorline, M. Schenk, P. Dallavalle, W. Shaffer, P. Davidson, and N. Seaman, 2004: Development and evaluation of the NOAA/EPA prototype air quality model prediction system. Available on the Internet at http://ams.confex.com/ams/pdfpapers/73739.pdf.

Peters A., D.W. Dockery, J.E. Muller, and M.A. Mittleman, 2001: Increased particulate air pollution and the triggering of myocardial infarction. Prepared for the American Heart Association. Available on the Internet at http://circ.ahajournals.org/cgi/content/abstract/103/23/2810.

U.S. Environmental Protection Agency (2003) Guidelines for developing an air quality (ozone and $PM_{2.5}$) forecasting program. Prepared by Dye T.S., MacDonald C.P., Anderson C.B., Hafner H.R., Wheeler N.J.M., and Chan A.C. at Sonoma Technology, Inc., Petaluma, CA, STI-902461-2295-FR, EPA-456/R-03-002, June. Available on the Internet at <http://www.airnow.gov/aq_forecasting_guidance-1016.pdf> ; last accessed October 28, 2004.

Venkatesh S., J.A. Pudykiewicz, A. Kallaur, R. Moffet, V.S. Bouchet, M. Jean, P.A. Makar, M.D. Moran, and W. Gong, 2003: Operational air quality forecasting in Canada: numerical model guidance for ground-level ozone and particulate matter. Available on the Internet at http://ams.confex.com/ams/pdfpapers/54490.pdf.

Wayland R.A., J.E.White, P.G. Dickerson, and T. Dye, 2002: Communicating real-time and forecasted air quality to the public: current state and future plans. *EM* (December 2002), 28-36.

Wayland R.A., J.E. White, T.S. Dye, C.B. Anderson, A.C. Chan, and D.E. Strohm, 2004: Future of AIRNow forecasting and the Air Quality Index: beyond mapping and forecasting. Paper presented at the *2004 American Meteorological Society Annual Meeting, AMS Sixth Conference on Atmospheric Chemistry: Air Quality in Megacities, Seattle, WA, January 11-15,* by the U.S. Environmental Protection Agency, Research Triangle Park, NC, and Sonoma Technology, Inc., Petaluma, CA (STI-902107-2406).

Acknowledgments. The authors thank Mr. Bill Murphey at the Georgia Department of Natural Resources, Mr. William Ryan at The Pennsylvania State University, and Mr. Daniel Salkovitz at the Virginia Department of Environmental Quality for providing forecast statistics data. We would also like to thank Dr. Jeffrey McQueen at NOAA for supplying gridded ozone forecast data. Special thanks goes to Mr. Richard Moffett from the Canadian Meteorological Center for providing CHRONOS data.

Table 1. Forecast statistics for Moderate AQI and above ozone days.

Forecast Statistics	**ATLANTA**			**BOSTON**			**BALTIMORE**		
	Human	NOAA	Statistical	Human	NOAA	Statistical	Human	NOAA	Statistical
Count	148	142	142	130	142	NA	147	142	133
PC (%)	79	70	74	90	85	NA	82	70	68
FAR (%)	31	44	39	53	68	NA	32	46	10
POD (%)	82	88	77	75	67	NA	90	73	90
CSI	60	52	51	41	28	NA	63	45	51
Bias	1	2	1	2	2	NA	1	1	1
Forecast Statistics	**CHARLOTTE**			**COLUMBUS**			**RICHMOND**		
	Human	NOAA	Statistical	Human	NOAA	Statistical	Human	NOAA	Statistical
Count	147	142	NA	152	142	144	136	142	132
PC (%)	80	71	NA	75	71	76	76	70	77
FAR (%)	33	46	NA	52	59	50	51	59	53
POD (%)	85	87	NA	64	63	40	77	93	74
CSI	60	50	NA	38	33	29	43	39	40
Bias	1	2	NA	1	2	1	2	2	2

Table 2. Forecast statistics for USG AQI and above ozone days.

Forecast Statistics	ATLANTA			BOSTON			BALTIMORE		
	Human	NOAA	Statistical	Human	NOAA	Statistical	Human	NOAA	Statistical
Count	148	142	142	130	142	NA	147	142	133
PC (%)	92	86	92	99	97	NA	95	92	95
FAR (%)	53	68	57	0	100	NA	33	60	20
POD (%)	64	90	60	0	0	NA	40	40	80
CSI	37	31	33	0	0	NA	33	25	36
Bias	1	3	1	0	3	NA	1	1	1
Forecast Statistics	**CHARLOTTE**			**COLUMBUS**			**RICHMOND**		
	Human	NOAA	Statistical	Human	NOAA	Statistical	Human	NOAA	Statistical
Count	147	142	NA	152	142	144	136	142	132
PC (%)	96	89	NA	97	89	98	99	96	97
FAR (%)	56	88	NA	100	88	100	100	83	80
POD (%)	80	50	NA	0	50	0	0	100	100
CSI	40	11	NA	0	11	0	0	17	20
Bias	2	4	NA	3	4	2	1	6	5

Table 3. Average ozone forecast statistics.

Average Statistics	Moderate AQI and above			Unhealthy for Sensitive Groups AQI and above		
	Human	NOAA	Statistical	Human	NOAA	Statistical
PC (%)	80	73	70	96	93	95
FAR (%)	42	54	44	57	83	78
POD (%)	79	78	73	31	47	36
CSI	51	41	41	18	14	13
Bias	1	2	1	1	3	2