EPA-454/R-96-004

Compilation of Photochemical Models' Performance Statistics for 11/94 Ozone SIP Applications

July 1996

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Emissions, Monitoring, and Analysis Division
Research Triangle Park, North Carolina 27711

U.S. Environmental Protection Agency
Region 5, Library (PL-12J)
77 West Jackson Boulevard, 12th Floor
Chicago, IL 60604-3590

DISCLAIMER

The information of this report has been reviewed in the entirety by the U.S. Environmental Protection Agency (EPA), and approved for publication as an EPA document. Mention of trade names, products, or services does not convey, and should not be interpretated as conveying official EPA approval, endorsement, or recommendation.

ACKNOWLEDGMENT

This report has been prepared by Sonoma Technology Incorporated and funded by the U. S. Environmental Protection Agency under Contract No. 68D30020 with Shao-Hang Chu as the Work Assignment Manager.

TABLE OF CONTENTS

Section		Page
ACKNOWLEDGMENT		iii
LIST OF FIGURES		vi
LIST OF TABLES		vii
1 INTRODUCTION		1-1
2. THE PHOTOCHEMICAL MODEL SIMULATIONS		2-1
3 MODEL PERFORMANCE		3-1
3.1	INDIVIDUAL SIMULATION STATISTICS	3-2
3.2	SIMULATION STATISTICS AVERAGED FOR EACH REGION	3-15
3.3	SIMULATION STATISTICS ON THE HIGHEST OZONE DAY IN EACH REGION	3-20
3.4	SIMULATION STATISTICS STRATIFIED BY MODELING METHODOLOGY AND GEOGRAPHIC SETTING	3-20
3.5	SIMULATION STATISTICS FOR THE UAM-IV, UAM-V, AND SAQM MODELS	3-31
4 CONCLUSIONS AND RECOMMENDATIONS		4-1
4.1	CONCLUSIONS	4-1
4.2	RECOMMENDATIONS	4-2
5 REFERENCES		5-1

LIST OF FIGURES

Figure		Page
3-1.	The predicted and observed domain-wide maximum ozone concentrations for SIP applications	3-10
3-2.	The number of simulations in each region with performance that met all three of EPA's ozone performance goals	3-13
3-3	The percentage of simulations in each region with performance that met all three of EPA's ozone performance goals	3-14
3-4	The average accuracy of the domain-wide peak ozone unpaired in space and time in each region	3-17
3-5.	The average normalized bias in predicted ozone concentrations above 60 ppb in each region	3-18
3-6	The average normalized gross error in predicted ozone concentrations above 60 ppb in each region	3-19
3-7	The accuracy of the domain-wide peak ozone unpaired in space and time on the highest ozone day in each region	3-22
3-8	The mean normalized bias in predicted ozone concentrations above 60 ppb on the highest ozone day in each region	3-23
3-9	The mean normalized gross error in predicted ozone concentrations above 60 ppb on the highest ozone day in each region	3-24
3-10	The predicted and observed domain-wide maximum ozone concentrations stratified by boundary condition methodology	3-28
3-11	The predicted and observed domain-wide maximum ozone concentrations stratified by wind model type	3-29
3-12	The predicted and observed domain-wide maximum ozone concentrations stratified by geophysical characteristics	3-30
3-13	The predicted and observed domain-wide maximum ozone concentrations stratified by the predominance of pollutant transport into the region	3-32

LIST OF TABLES

Table		Page
2-1.	Regions included in the study	2-2
3-1.	The photochemical models' ozone performance statistics for individual baseline simulations	3-3
3-2.	The extent to which EPA performance goals were achieved in the SIP applications	3-11
3-3.	Average photochemical models' ozone performance statistics by region	3-16
3-4	Photochemical models' ozone performance statistic on high ozone day by region	3-21
3-5	The methodologies used to develop boundary conditions and windfields, and the characteristics of the geographic setting	3-25
3-6	Photochemical models' performance statistics for ozone stratified by boundary condition methodology, windfield model, geographic setting, and predominance of pollutant transport into the region	3-27
3-7	Comparison of the average model performance statistics for ozone from the SAQM, UAM-IV, and UAM-V models	3-31
4-1	Additional statistical parameters for model performance evaluation	4-3

LIST OF TABLES

Table		Page
2-1.	Regions included in the study	2-2
3-1.	The photochemical models' ozone performance statistics for individual baseline simulations	3-3
3-2.	The extent to which EPA performance goals were achieved in the SIP applications	3-11
3-3.	Average photochemical models' ozone performance statistics by region	3-16
3-4.	Photochemical models' ozone performance statistic on high ozone day by region	3-21
3-5.	The methodologies used to develop boundary conditions and windfields, and the characteristics of the geographic setting	3-25
3-6.	Photochemical models' performance statistics for ozone stratified by boundary condition methodology, windfield model, geographic setting, and predominance of pollutant transport into the region	3-27
3-7.	Comparison of the average model performance statistics for ozone from the SAQM, UAM-IV, and UAM-V models	3-31
4-1	Additional statistical parameters for model performance evaluation	4-3

1. INTRODUCTION

The scientific credibility of photochemical modeling studies depends on the soundness of the model formulation, the adequacy of the aerometric database, and the accuracy of the model's predictions for specific applications (Tesche et al., 1990; Roth et al., 1991). The accuracy of specific applications must be evaluated because the quality and representativeness of input data can strongly influence the model performance. An investigation was carried out to evaluate the performance of three grid-based photochemical models in recent applications. The photochemical models were the Urban Airshed Model Version IV (UAM-IV), the Urban Airshed Model variable grid version (UAM-V), and the San Joaquin Valley Air Quality Model (SAQM). All three models employed the CB4 chemical mechanism for these applications. The models were applied in 24 ozone nonattainment regions in 1993-1995 to support the development of emissions control strategies for the 1994 ozone State Implementation Plans (SIPs). The models were used to simulate the relationship between nitrogen oxides (NO_x) and volatile organic compound (VOC) emissions, and ambient ozone concentrations under current episodic conditions and for future emission scenarios. The baseline model applications provide an unique opportunity to examine photochemical models' performance for a large number of regions and episodes. Furthermore, the SIP simulations were performed using generally consistent methodologies.

The U.S. Environmental Protection Agency (EPA) established model performance goals for the SIP applications. In order to assess the extent to which the baseline simulations met the performance goals, the EPA required states to report three statistical measures of the model's ability to predict ambient ozone concentrations. These measures are:

1. The normalized accuracy of domain-wide maximum 1-hr concentration unpaired in space and time.

2. Mean normalized bias of all predicted and observed concentration pairs where the observed concentrations exceed 60 ppb.

3. Mean normalized error of all predicted and observed concentration pairs where the observed concentrations exceed 60 ppb.

This report reflects the status of November 1994 SIP applications model performance results. As the States continue their modeling efforts these results are expected to change. This document is a compilation of the EPA recommended basic model performance statistics for the 24 ozone nonattainment areas in the November 1994 SIP applications. For a more complete evaluation of the models' performance both spatial and temporal analyses of the matching between the predicted and observed concentrations are desirable. Thus, it is important to recognize that the scope of this evaluation is quite limited. The evaluation focuses on three statistical measures of the models' ability to predict ambient ozone concentrations. It does not include an evaluation of the models' performance on other species (such as NO, NO_2, NO_y, and VOCs) or for other statistical measures of its performance on ozone, which may be important for establishing the suitability of simulations for use in control strategy development.

2. THE PHOTOCHEMICAL MODEL SIMULATIONS

The statistical performance data for this study were obtained from reports and briefings submitted to the EPA. Model performance results for the 22 regions listed in **Table 2-1** were included in the database. In some cases, more than one nonattainment area was included in a single modeling domain and separate statistics were included for each area if they were provided in the modeling documentation (e.g., in the Houston and Beaumont, Texas areas). In most cases the reported statistics apply to a single nonattainment area. Data for several other regions were received but not included as explained below.

The photochemical model simulations were carried out by numerous modeling groups under the direction of state and local agencies. EPA provided modeling guidelines (EPA, 1991) to promote consistency between the applications. The guidelines allowed considerable flexibility so that regions with different types of databases, different geographic settings and meteorological conditions, and different levels of modeling expertise could complete the applications. As shown in Table 2-1, most of the simulations were performed with the UAM-IV model (SAI, 1990), version 6.21. However, simulations for the Lake Michigan Ozone Study (LMOS) region (i.e., the four states surrounding Lake Michigan) were performed with the UAM-V model (SAI, 1993, 1994a). Simulations for the greater Atlanta, Georgia region were carried out with the UAM-IV and UAM-V models. Comparable efforts were made in the applications of both models in Atlanta which facilitates direct comparison of the results. Differences in the predictions from the UAM-IV and UAM-V models are small for Atlanta and the UAM-IV results were used in the summary of model performance. A comparison of performance statistics for the Atlanta UAM-IV and UAM-V simulations are included. In addition, the SAQM model was used for California's San Joaquin Valley modeling domain. The SAQM model is an enhanced version of the Regional Acid Deposition Model (RADM).

The photochemical simulations were performed for two- to seven-day periods with one or more days with 1-hr ozone concentrations above 120 ppb. The results for the first day of each period were excluded from statistical analysis because it takes time for the simulations to become driven by emissions rather than initial concentration estimates. In some large modeling domains (e.g., San Joaquin Valley), both the first and second days of the simulations were treated as "start-up" days. Table 2-1 shows the number of episode days modeled in each region. Only modeled episode days with observed ozone concentrations above 120 ppb were included in the database. On average, six episode days were modeled in each region. The number of days modeled varied significantly between the regions. For example, 17 days were modeled in the New York area and only one day was modeled in the Sacramento area. The number of episode days modeled was generally higher in the areas with more severe ozone problems (i.e., in Los Angeles, Houston, and New York). The total number of episode days included in the database is 131 days.

Table 2-1. Regions included in the study.

Region	Model	Number of Episode Days Simulated	Reference
San Diego, CA	UAM-IV	2	SDAPCD, 1994
Los Angeles, CA	UAM-IV	10	SCAQMD, 1994
Ventura, CA	UAM-IV	3	CARB, 1995a
San Joaquin Valley, CA	SAQM	2	CARB, 1995c
Sacramento, CA	UAM-IV	1	CARB, 1995b
Phoenix, AZ	UAM-IV	2	Briefing
Houston, TX	UAM-IV	12	TNRCC, 1994b
Beaumont, TX	UAM-IV	10	TNRCC, 1994b
Dallas TX	UAM-IV	9	TNRCC, 1994c
Baton Rouge, LA	UAM-IV	7	SAI, 1994b
St Louis MO	UAM-IV	8	Briefing
Lake Michigan	UAM-V	8	LADCO, 1994 & Briefing
Detroit, MI	UAM-IV	4	Briefing
Nashville TN	UAM-IV	4	Kaminski, 1995
Louisville, KY	UAM-IV	3	Briefing
Cincinnati, OH	UAM-IV	6	Briefing
Atlanta GA	UAM-IV	4	SAI, 1994c
Richmond VA	UAM-IV	6	Briefing
Philadelphia, NJ	UAM-IV	5	Georgopoulos, 1995
New York, NY	UAM-IV	17	NYDEC, 1994a,b,c,d
Baltimore, MD	UAM-IV	4	Briefing
New England	UAM-IV	4	Briefing

As noted above, the database does not include results for all of the ozone nonattainment areas The UAM-IV model was applied in El Paso, Texas without emissions for a significant portion of the modeling domain (Juarez, Mexico) and the results showed gross underprediction of observed ozone concentrations (TNRCC, 1994a). The El Paso simulation results were excluded because they were based on grossly inadequate inputs compared to all other simulations

considered here. Results for the Santa Barbara and Southeast Desert Air Basin of California were not reported separately from those for Ventura and the South Coast Air Basin of California, respectively.

The techniques employed to prepare the meteorological and air quality model inputs varied between the regions. Numerous modelers used prognostic meteorological models, such as the MM5, CSUMM, and CALRAMS, to develop the hourly 3-dimensional wind fields, while other modelers chose diagnostic wind models, such as CALMET, UAM-Diagnostic Wind Model, and the Regional Oxidant Model (ROM) meteorological model. Various models were also used to develop mixing height (RAMMET, MIXEMUP, CALMET, etc.). Likewise, boundary concentration inputs were obtained from the EPA- recommended default values, surface observations, surface and aircraft observations, and regional model estimates (ROM). The difference in model input preparation procedures were often determined by the extent of the aerometric database in the regions and the complexity of the meteorology. Prognostic meteorological models were used more frequently in areas with coastal meteorology and/or complex terrain (e.g., Los Angeles and LMOS) rather than flat areas with continental meteorology. To the extent possible, the database was coded with the model input preparation procedure codes.

The photochemical model performance statistics incorporated into the database were those provided by the modeling groups in reports and briefings to the EPA. STI did not recompute the statistics from the model output files and the observed concentration data because they were not made available for use in the study. The absence of actual model output and observation files inhibited quality assuring the statistical data and limited the scope of the analysis to those statistics reported by the modeling groups. While numerous modeling groups reported more than the required statistics, only the three EPA-required statistics were common to all of the application reports.

3. MODEL PERFORMANCE

This evaluation of model performance in SIP applications focuses on the models' ability to predict the domain-wide peak ozone concentration and the concentrations at all locations with observed ozone concentrations above 60 ppb. The three statistical measures recommended in the EPA photochemical modeling guidelines (EPA, 1991) are used. The measures are:

1. The normalized accuracy of domain-wide maximum 1-hr concentration unpaired in space and time (A_u):

$$A_u = 100\left(\frac{Pred_{domain\text{-}wide\ peak} - Obs_{domain\text{-}wide\ peak}}{Obs_{domain\text{-}wide\ peak}}\right) \quad (1)$$

2. Mean normalized bias of all predicted and observed concentration pairs where the observed concentrations exceed 60 ppb ($NBIAS_{60}$):

$$NBIAS_{60} = \frac{100}{N}\sum_{i=1}^{N}\frac{(Pred_{xi} - OBS_{xi})}{Obs_{xi}} \quad (2)$$

where N includes all of the predicted and observed concentratio pairs with observed concentrations above 60 ppb.

3. Mean normalized error of all predicted and observed concentration pairs where the observed concentrations exceed 60 ppb ($NERROR_{60}$):

$$NERROR_{60} = \frac{100}{N}\sum_{i=1}^{N}\frac{Pred_{xi} - OBS_{xi}}{Obs_{xi}} \quad (3)$$

where N includes all of the predicted and observed concentratio pairs with observed concentrations above 60 ppb

Note the sign convention of the peak accuracy and normalized bias as defined above are the opposite of that recommended in the EPA guidelines. This sign convention is used here because 19 of the 22 regions reported their bias results using this convention. That is, they defined model overprediction as positive bias, rather than as negative bias. Because STI did not have access to the model output and observed data files to recompute the statistics in a consistent manner, the most common reporting convention was employed. The statistics for the three regions (Nashville, Richmond, and Atlanta), which reported using the EPA sign convention, were converted to the common convention by changing the sign on the reported mean bias statistics. The unpaired peak accuracy statistics were computed from the observed and predicted data.

The EPA guidelines set statistical performance goals for regulatory applications of the UAM. The goal for the peak accuracy unpaired in space and time is within ±15 percent. The goal for the mean normalized bias is within ±15 percent. The goal for the mean normalized error is less than 35 percent. These goals are primarily based on historical performance of the model, rather than analyses of accuracy requirements for intended use of the model (e.g., analyses that relate accuracy of baseline simulations to accuracy of the response of the model to emission changes).

3.1 INDIVIDUAL SIMULATION STATISTICS

The statistical performance data for the 129 individual simulations are shown in **Table 3-1** and **Figure 3-1**. The table indicates the region, the date of the episode day, the model, the observed domain-wide maximum 1-hr ozone concentration, the predicted domain-wide maximum 1-hr ozone concentration, the accuracy of the domain-wide peak, the mean normalized bias, and the mean normalized error. The average, standard deviations, minimum, and maximum values are shown at the bottom of the table. **Table 3-2** and **Figures 3-2 and 3-3** illustrate the extent to which the individual simulations met EPA's performance goals.

The results indicate that the models predicted an average domain-wide peak of 172 ±51 ppb when the average observed domain-wide peak was 170 ±44 ppb. On average, the accuracy of the predicted domain-wide peak ozone concentration unpaired in space and time was +2 ± 20 percent. The near absence of bias on average in the unpaired peak ozone is an excellent result for air quality models. Even at the extremes, the ozone peak accuracy is only a factor of two in error. The lowest accuracy domain-wide peaks in individual simulations were for October 12, 1991 in Houston where the ozone was underpredicted by 53 percent (112 ppb predicted versus 240 ppb observed) and for June 21, 1988 in New York where the ozone was overpredicted by 100 percent (298 ppb predicted versus 149 ppb observed). At least one simulation in every region met the ±15 percent performance goal for the unpaired peak accuracy. Overall, 77 percent of the individual simulations met this performance goal. For areas that modeled four or more episode days, over 75 percent of the episode days modeled in the following regions met the EPA performance goal for unpaired peak ozone: Los Angeles, Dallas, St. Louis, Lake Michigan, Detroit, Atlanta, and Baltimore. The statistics and Figure 3-1 show there is considerable scatter in the unpaired peak predictions. The correlation between the observed and predicted domain-wide peak is moderate, $R^2 = 0.54$, which indicates the models are able to explain 54 percent of the variance in the peak ozone.

The mean normalized bias for ozone concentration above 60 ppb averaged -5 ± 16 percent in the SIP applications. The negative bias is consistent with the majority of previous UAM simulations. The small average bias is comparable to the uncertainty in the ozone observations (±5 ppb). This level of performance on average is excellent for air quality models. At the extremes, the mean normalized bias was as low as -45 percent (October 12, 1991 in Houston) and as high as +55 percent (June 21, 1998 in New York) in individual simulations. At least one simulation in every region met the ±15 percent performance goal for the mean domain-wide peak ozone was overpredicted by 6.4 percent on average with the ROM-derived boundary

Table 3-1. The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias[a] (%)	Mean Gross Error[a] (%)
San Diego, CA	UAM IV	August 29, 1989	154	141	-8.4	-12.9	24.0
San Diego, CA	UAM-IV	September 21, 1989	156	135	-13.5	-5.6	31.0
Los Angeles, CA[b]	UAM-IV	August 27, 1987	240	216	-10.0	-22.0	30.0
Los Angeles, CA[b]	UAM-IV	August 28, 1987	290	322	11.0	-8.0	20.0
Los Angeles, CA[b]	UAM-IV	July 14, 1987	250	255	2.0	-24.0	32.0
Los Angeles, CA[b]	UAM IV	July 15, 1987	220	288	31.0	-12.0	27.0
Los Angeles, CA[b]	UAM-IV	June 24, 1987	230	251	9.0	-25.0	33.0
Los Angeles, CA[b]	UAM-IV	June 25, 1987	240	250	4.0	-26.0	36.0
Los Angeles, CA[b]	UAM-IV	June 6, 1985	360	392	9.0	7.0	33.0
Los Angeles, CA[b]	UAM-IV	June 7, 1985	360	382	6.0	7.0	32.0
Los Angeles, CA[b]	UAM-IV	September 8, 1987	330	254	-23.0	-27.0	34.0
Los Angeles, CA[b]	UAM-IV	September 9, 1987	260	252	-3.0	-21.0	36.0
Ventura, CA	UAM-IV	September 17, 1984	140	143	2.1	-12.0	24.0
Ventura, CA	UAM-IV	September 6, 1984	170	175	2.9	-12.0	20.0
Ventura, CA	UAM-IV	September 7, 1984	180	173	-3.9	-2.0	21.0
San Joaquin Valley, CA	SAQM	August 5, 1990	150	139	-7.3	1.0	16.0
San Joaquin Valley, CA	SAQM	August 6, 1990	160	150	-6.3	-7.0	15.0
Sacramento, CA	UAM-IV	July 13, 1990	143	139	-2.8	-12.0	18.0
Phoenix, AZ	UAM-IV	August 10, 1992	157	134	-14.6	-5.4	23.9
Phoenix, AZ	UAM-IV	June 14, 1993	140	132	-5.7	4.4	20.4
Houston, TX	UAM IV	August 1, 1990	138	163	18.1	-40.8	42.5

Table 3.1. The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain-wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias* (%)	Mean Gross Error* (%)
Houston, TX	UAM-IV	July 27, 1990	180	185	2.8	-25.7	30.1
Houston, TX	UAM-IV	July 28, 1990	150	179	19.3	-31.3	34.6
Houston, TX	UAM-IV	July 29, 1990	203	263	29.6	-14.1	32.2
Houston, TX	UAM-IV	July 30, 1990	191	253	32.5	17.7	32.6
Houston, TX	UAM-IV	July 31, 1990	150	211	40.7	-11.4	43.9
Houston, TX	UAM-IV	May 16, 1988	160	139	-13.1	-32.9	36.0
Houston, TX	UAM-IV	May 17, 1988	210	153	-27.1	-28.6	38.0
Houston, TX	UAM-IV	May 18, 1988	210	173	-17.6	-36.3	38.2
Houston, TX	UAM-IV	October 10, 1991	122	151	23.8	-8.9	21.0
Houston, TX	UAM-IV	October 11, 1991	189	131	-30.7	-5.4	21.7
Houston, TX	UAM-IV	October 12, 1991	240	112	-53.3	-45.0	45.2
Beaumont, TX	UAM-IV	July 28, 1990	129	146	13.2	10.5	19.5
Beaumont, TX	UAM-IV	July 30, 1990	139	174	25.2	23.1	32.5
Beaumont, TX	UAM-IV	July 31, 1990	226	151	-33.2	-31.8	32.0
Beaumont, TX	UAM-IV	May 18, 1988	130	163	25.4	-13.9	16.6
Beaumont, TX	UAM-IV	May 19, 1988	180	139	-22.8	-27.5	29.9
Beaumont, TX	UAM-IV	May 19, 1988	180	139	-22.8	-27.5	29.9
Beaumont, TX	UAM-IV	October 10, 1991	144	109	-24.3	-6.6	17.3
Beaumont, TX	UAM-IV	October 11, 1991	144	127	-11.8	-5.7	14.1
Beaumont, TX	UAM-IV	October 12, 1991	200	115	-42.5	-42.8	43.2
Beaumont, TX	UAM-IV	October 13, 1991	137	121	-11.7	-34.2	34.2

Table 3-1. The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain-wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias* (%)	Mean Gross Error* (%)
Dallas, TX	UAM-IV	August 1, 1991	170	155	-8.8	-5.7	23.1
Dallas, TX	UAM-IV	August 25, 1988	130	131	0.8	-6.9	17.1
Dallas, TX	UAM-IV	August 26, 1988	160	144	-10.0	-10.1	20.4
Dallas, TX	UAM-IV	August 27, 1990	140	147	5.0	3.8	18.4
Dallas, TX	UAM-IV	August 28, 1990	160	133	-16.9	-9.6	18.8
Dallas, TX	UAM-IV	August 29, 1990	140	146	4.3	-1.7	13.0
Dallas, TX	UAM-IV	August 30, 1990	160	145	-9.4	-6.2	16.4
Dallas, TX	UAM-IV	July 31, 1991	130	133	2.3	-16.4	28.9
Dallas, TX	UAM-IV	June 18, 1987	160	169	5.6	4.9	13.8
Baton Rouge, LA	UAM-IV	August 15, 1989	135	159	17.8	-8.1	22.2
Baton Rouge, LA	UAM-IV	August 16, 1989	157	160	1.9	1.3	21.2
Baton Rouge, LA	UAM-IV	August 21, 1992	135	158	17.0	-18.1	23.5
Baton Rouge, LA	UAM-IV	July 27, 1989	130	130	0.0	-1.3	22.1
Baton Rouge, LA	UAM-IV	July 28, 1989	129	140	8.5	-21.1	32.1
Baton Rouge, LA	UAM-IV	May 24, 1990	123	138	12.2	3.6	12.1
Baton Rouge, LA	UAM-IV	May 25, 1990	152	164	7.9	9.9	20.0
St. Louis, MO	UAM-IV	August 15, 1988	152	142	-6.6	14.0	23.0
St. Louis, MO	UAM-IV	August 16, 1988	142	145	2.1	7.0	21.0
St. Louis, MO	UAM-IV	August 17, 1988	157	146	-7.0	-9.0	23.0
St. Louis, MO	UAM-IV	August 18, 1988	149	123	-17.4	-16.0	25.0
St. Louis, MO	UAM-IV	July 7, 1988	125	138	10.4	1.0	14.0

Table 3-1. The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain-wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias[a] (%)	Mean Gross Error[a] (%)
St Louis, MO	UAM-IV	July 8, 1988	163	160	-1.8	-7.0	20.0
St Louis, MO	UAM-IV	July 9, 1988	126	140	11.1	-8.0	16.0
St Louis, MO	UAM-IV	June 24, 1987	160	152	-5.0	7.0	24.0
Lake Michigan	UAM-V	August 26, 1991	178	150	-16.0	5.0	14.0
Lake Michigan	UAM-V	July 17, 1991	139	142	2.0	11.0	17.0
Lake Michigan	UAM-V	July 18, 1991	169	161	-5.0	4.0	16.0
Lake Michigan	UAM-V	July 19, 1991	153	144	-6.0	13.0	18.0
Lake Michigan	UAM-V	June 20, 1991	121	109	-10.0	-7.0	14.0
Lake Michigan	UAM-V	June 21, 1991	134	151	13.0	-12.0	20.0
Lake Michigan	UAM-V	June 26, 1991	175	165	-6.0	3.0	14.0
Lake Michigan	UAM-V	June 28, 1991	133	137	3.0	10.0	16.0
Detroit, MI	UAM-IV	August 2, 1988	144	142	-1.4	-12.0	22
Detroit, MI	UAM-IV	August 3, 1988	164	148	-9.8	-24.0	28
Detroit, MI	UAM-IV	July 5, 1988	147	128	-12.9	-15	20
Detroit, MI	UAM-IV	July 7, 1988	168	164	-2.4	-12.0	22
Louisville, KY	UAM-IV	July 20, 1987	151	187	23.8	1.7	15.9
Louisville, KY	UAM-IV	June 23, 1988	137	138	0.7	-28.1	29.5
Louisville, KY	UAM-IV	June 27, 1990	148	160	8.1	6.0	16.1
Nashville, TN	UAM-IV	August 1, 1988	138	115	-16.7	-7.2[c]	15.1
Nashville, TN	UAM-IV	August 3, 1988	135	132	-2.2	-9.4[c]	20.2
Nashville, TN	UAM-IV	July 9, 1988	138	139	0.7	-9.0[c]	20.3

Table 3-1 The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain wide Peak Unpaired (%)	Mean Normalized Bias[a] (%)	Mean Gross Error[a] (%)
Nashville, TN	UAM IV	June 24, 1988	137	113	-17.5	-25.2[c]	26.1
Cincinnati, OH	UAM-IV	August 1, 1988	155	190	22.7	-4.8	22.5
Cincinnati, OH	UAM-IV	August 16, 1988	144	158	9.9	-3.5	19.2
Cincinnati, OH	UAM-IV	August 17, 1988	144	150	4.0	-14.6	21.7
Cincinnati, OH	UAM-IV	August 18, 1988	159	150	-5.6	-10.0	21.6
Cincinnati, OH	UAM-IV	August 2, 1988	140	180	28.3	-3.3	20.4
Cincinnati, OH	UAM-IV	August 3, 1988	169	169	0.0	3.5	19.4
Atlanta, GA	UAM-IV	July 31, 1987	201	194	-3.4	5.2[c]	28
Atlanta, GA	UAM-IV	August 1, 1987	169	191	12.9	-0.9[c]	13.2
Atlanta, GA	UAM-IV	July 8, 1988	186	197	6.1	-8.5[c]	18.4
Atlanta, GA	UAM-IV	Aug. 10, 1992	132	148	11.9	4.3[c]	22.1
Richmond, VA	UAM-IV	July 10, 1988	142	127	-10.6	18.2[c]	20.7
Richmond, VA	UAM-IV	July 15, 1988	141	142	0.7	-8.7[c]	13.3
Richmond, VA	UAM-IV	July 6, 1988	155	130	-16.1	-15.0[c]	20.6
Richmond, VA	UAM-IV	July 7, 1988	173	146	-15.6	-15.9[c]	18.8
Richmond, VA	UAM-IV	June 13, 1988	144	131	-9.0	-12.3[c]	18.2
Richmond, VA	UAM-IV	June 14, 1988	153	144	-5.9	-7.4[c]	19.8
Philadelphia, NJ	UAM-IV	July 19, 1991	150	160	6.7	24.3	35.0
Philadelphia, NJ	UAM-IV	July 20, 1991	180	304	68.8	15.3	29.6
Philadelphia, NJ	UAM-IV	July 7, 1988	210	201	-4.3	6.9	25.6
Philadelphia, NJ	UAM-IV	July 8, 1988	210	274	30.4	20.1	30.7

Table 3-1. The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain-wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias* (%)	Mean Gross Error* (%)
Philadelphia, NJ	UAM-IV	June 15, 1987	190	211	11.0	16.1	30.4
New York, NY	UAM-IV	July 10, 1988	216	203	-6.0	3.2	24.3
New York, NY	UAM-IV	July 11, 1988	222	216	-2.7	17.7	27.9
New York, NY	UAM-IV	July 17, 1991	155	152	-1.9	-7.6	17.4
New York, NY	UAM-IV	July 18, 1991	174	188	8.0	3.6	17.3
New York, NY	UAM-IV	July 19, 1991	161	162	0.6	5.4	21.7
New York, NY	UAM-IV	July 20, 1991	175	245	40.0	2.6	20.2
New York, NY	UAM-IV	July 7, 1988	193	222	15.0	35.1	43.9
New York, NY	UAM-IV	July 8, 1988	212	219	3.3	18.8	29.1
New York, NY	UAM-IV	July 9, 1988	149	298	100.0	55.3	56.5
New York, NY	UAM-IV	June 15, 1987	196	202	3.1	17.3	25.2
New York, NY	UAM-IV	June 18, 1987	125	147	17.6	3.8	18.4
New York, NY	UAM-IV	June 19, 1987	167	172	3.0	19.6	27.5
New York, NY	UAM-IV	June 19, 1988	153	178	16.3	24.3	29.2
New York, NY	UAM-IV	June 20, 1987	294	206	-29.9	-2.9	28.7
New York, NY	UAM-IV	June 20, 1988	149	165	10.7	12.7	22.9
New York, NY	UAM-IV	June 21, 1988	134	235	75.4	14.4	25.9
New York, NY	UAM-IV	June 22, 1988	239	192	-19.7	2.4	21.9
Baltimore, MD	UAM-IV	June 29, 1988	180	169	-6.1	3.2	14.9
Baltimore, MD	UAM-IV	June 30, 1988	172	172	0.0	0.4	21.7
Baltimore, MD	UAM-IV	June 19, 1988	132	161	22.0	5.6	18.2

Table 3-1. The photochemical models' ozone performance statistics for individual baseline simulations.

Region	Model	Episode Date	Observed Domain wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias[a] (%)	Mean Gross Error[a] (%)
Baltimore, MD	UAM-IV	June 20, 1988	164	157	-4.3	-9.9	22.3
New England	UAM-IV	July 8, 1988	184	186.6	1.4	-9.9	25.8
New England	UAM-IV	July 11, 1988	222	210.7	-5.1	-12.1	24
Average		129.0	170	172	2	-5	24
Standard Deviation		129.0	44	51	20	16	8
Minimum		129.0	121	109	-53	-45	12
Maximum		129.0	360	392	100	55	57

[a] For all locations and hours with observed ozone concentrations above 60 ppb.

[b] The statistics for Los Angeles (or California's South Coast Air Basin) only included concentrations above 80 to 120 ppb, depending on the episode, rather than all concentrations above 60 ppb. Also, all of the Los Angeles results are for the simulations with double VOC emissions.

[c] Approximate values because the original data were reported with the opposite sign convention.

[d] The Philadelphia statistics are for the Case 4S simulations.

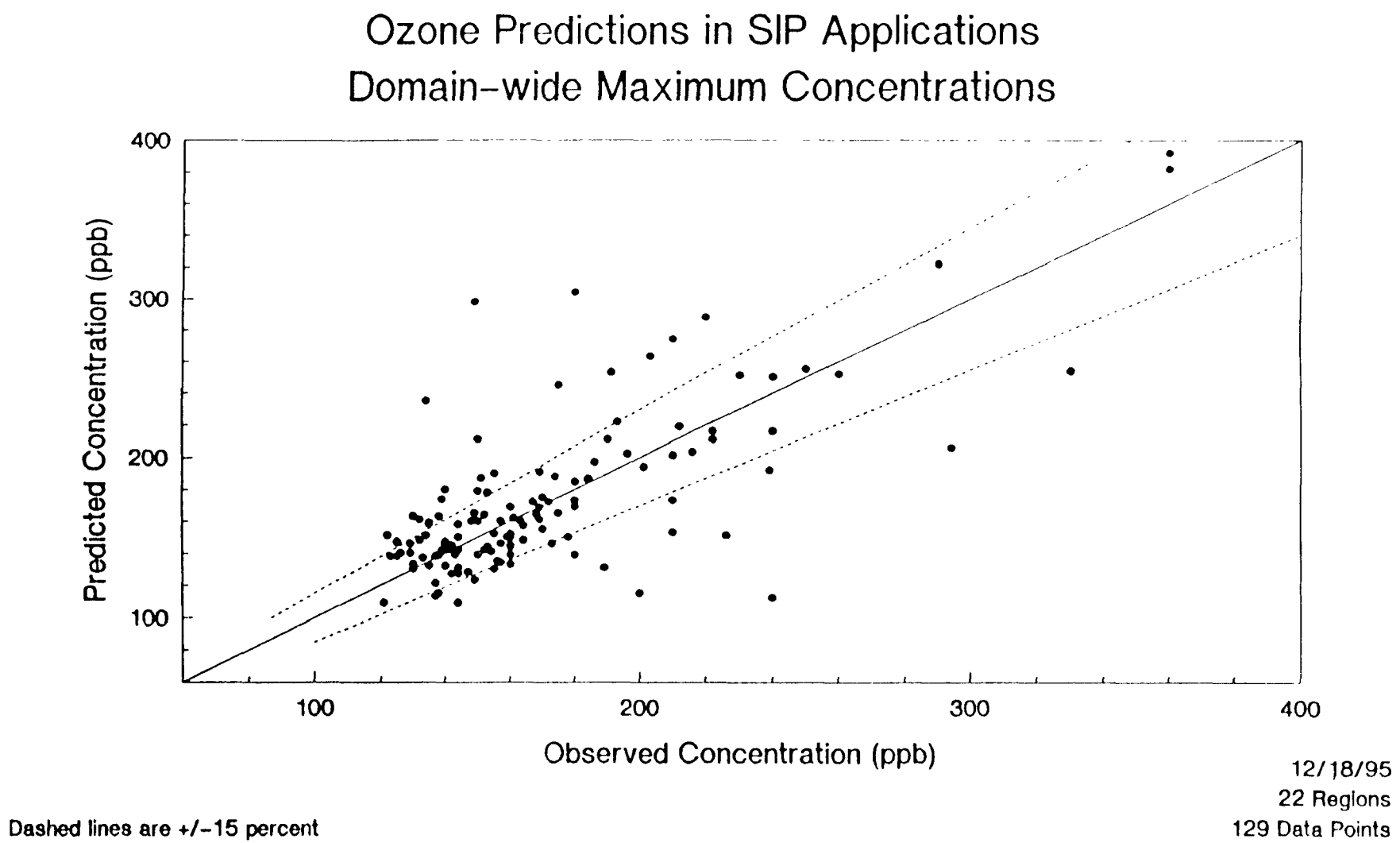

Figure 3-1. The predicted and observed domain-wide maximum ozone concentrations for SIP applications.

Table 3-2. The extent to which EPA performance goals were acheived in the SIP applications.

Region	Number of Days Modeled	Number of Days for Which Performance Goals Were Achieved				Percent of Modeled Days for Which Performance Goals Were Achieved			
		Peak Accuracy Criteria	Mean Bias Criteria	Mean Error Criteria	All Criteria	Peak Accuracy Criteria	Mean Bias Criteria	Mean Error Criteria	All Criteria
San Diego	2	2	2	2	2	100	100	100	100
Los Angeles	10	8	4	8	3	80	40	80	30
Ventura	3	3	3	3	3	100	100	100	100
San Joaquin Valley	2	2	2	2	2	100	100	100	100
Sacramento	1	1	1	1	1	100	100	100	100
Phoenix	2	2	2	2	2	100	100	100	100
Houston	12	2	4	6	0	17	33	50	0
Beaumont	10	3	4	9	2	30	40	90	20
Dallas	9	8	8	9	7	89	89	100	78
Baton Rouge	7	5	5	7	4	71	71	100	57
St. Louis	8	7	7	8	7	88	88	100	88
Lake Michigan	8	7	8	8	7	88	100	100	88
Detroit	4	4	2	4	2	100	50	100	50
Louisville	3	2	2	3	1	67	67	100	33
Nashville	4	2	3	4	2	50	75	100	50
Cincinnati	6	4	6	6	4	67	100	100	67
Atlanta	4	4	4	4	4	100	100	100	100
Richmond	6	4	3	6	3	67	50	100	50
Philadelphia	5	3	1	4	1	60	20	80	20

Table 3-2. The extent to which EPA performance goals were acheived in the SIP applications.

Region	Number of Days Modeled	Number of Days for Which Performance Goals Were Achieved				Percent of Modeled Days for Which Performance Goals Were Achieved			
		Peak Accuracy Criteria	Mean Bias Criteria	Mean Error Criteria	All Criteria	Peak Accuracy Criteria	Mean Bias Criteria	Mean Error Criteria	All Criteria
New York	17	9	10	15	5	53	59	88	29
Baltimore	4	3	4	4	3	75	100	100	75
New England	2	2	2	2	2	100	100	100	100
Sum		87	86	117	66				
Mean						77.3	75.3	94.9	64.1
Minimum		1	1	1	0	16.7	20.0	50.0	0.0
Maximum		9	10	15	7	100.0	100.0	100.0	100.0

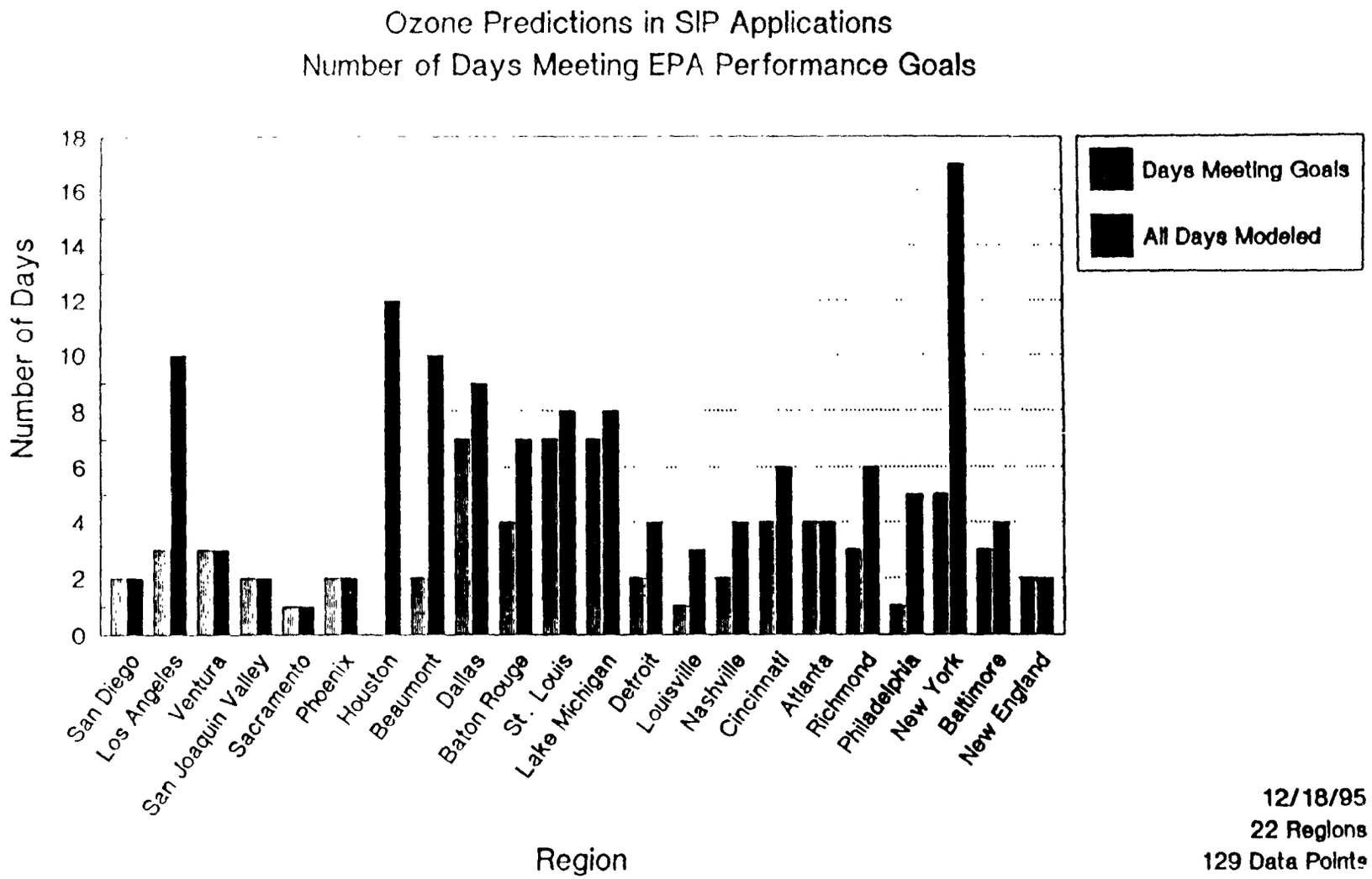

Figure 3-2. The number of simulations in each region with performance that met all three of EPA's ozone performance goals.

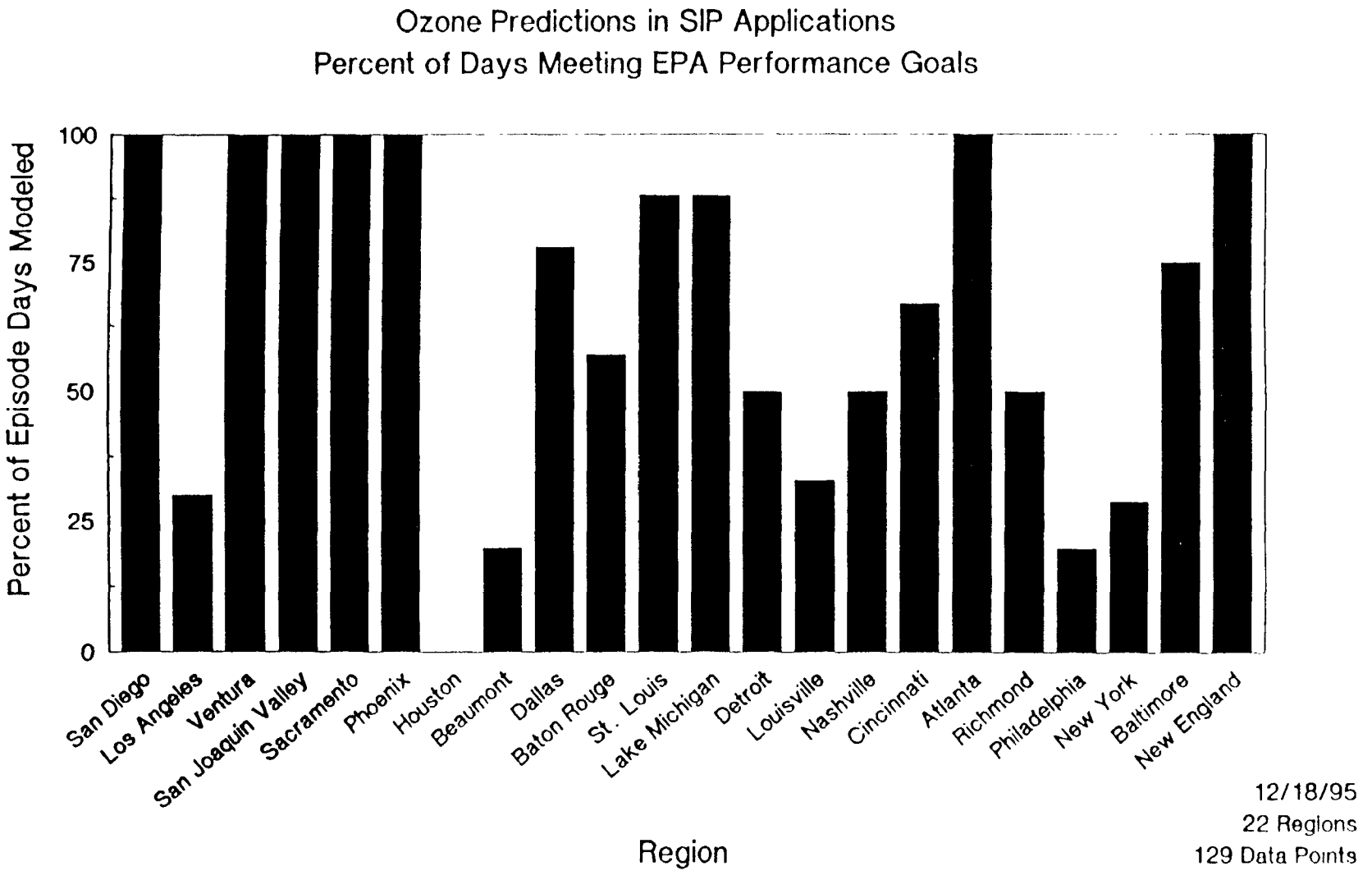

Figure 3-3. The percentage of simulations in each region with performance that met all three of EPA's ozone performance goals.

areas that modeled four or more episode days, over 75 percent of the episode days modeled in the following regions met the EPA performance goal for mean normalized bias: Dallas, St. Louis, Lake Michigan, Nashville, Cincinnati, Atlanta, and Baltimore.

The mean normalized gross error averaged 24 ± 8 percent in the SIP simulations. This level of error is somewhat lower than in most historical UAM applications. In individual simulations, the gross error was as low as 12 percent (May 24, 1990 in Baton Rouge) and as high as 56 percent (June 21, 1988 in New York). In most regions, all of the simulations had gross errors of less than 35 percent for ozone. However, selected simulations for Los Angeles, Houston, Beaumont, Richmond, and New York had higher gross error than the EPA goal. Overall, 95 percent of the simulations met the EPA goal for gross error.

Table 3-2 shows the number of simulations that met all three of the EPA's performance goals. At least one simulation in every region except Houston met all three goals. Overall, 64 percent of the simulations met all three goals. However, several of the areas that modeled a large number of days had low percentages of simulations achieving all three goals. For example, only 3 of 10 days in Los Angeles, 0 of 12 days in Houston, 2 of 10 in Beaumont, and 5 of 17 days in New York met all three goals. The simulations for Dallas, Lake Michigan, and St. Louis had the most days with performance that met all three goals.

3.2 SIMULATION STATISTICS AVERAGED FOR EACH REGION

Table 3-3 and **Figures 3-4 through 3-6** show the model performance statistics averaged for each region. They show the domain-wide peak ozone was underpredicted by more than 9 percent (1 σ) in San Diego, Phoenix, and Beaumont and overpredicted by more than 9 percent in Louisville, Cincinnati, Philadelphia, and New York on average. The domain-wide peak ozone was predicted within ±10 percent in the 15 other regions on average. The correlation between the region averaged predicted and observed peaks is high, $R^2 = 0.86$, and considerably better than that for the individual simulations.

The grand average mean bias averaged by region is -5.6 percent, which is similar to the grand average for all individual simulations. In Los Angeles, Houston, Beaumont, and Detroit, the average predictions have negative biases of 15 percent (1 σ) or more. Positive average biases of 13 percent (+2σ) or more are evident in the Philadelphia and New York simulations. The mean bias in the other 16 regions is between -13 and +3 percent, which is reasonably good.

The grand average mean gross error averaged by region is 22.7 percent which is slightly better than the average for all individual simulations. The results show that the mean gross error is generally higher (greater than 25 percent) in the region with high ambient ozone concentrations (Los Angeles, Houston, Philadelphia, New York, and New England). The simulations for San Joaquin Valley, Lake Michigan, Sacramento, Dallas, Richmond, and Baltimore have mean gross error of less than 20 percent on average, which is quite good.

Table 3-3. Average photochemical models' ozone performance statistics by region

Region	Number of Days	Observed Domain-wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias (%)	Mean Gross Error (%)
San Diego, CA	2	155	138	-11	-9	27
Los Angeles, CA	10	278	286	4	-15	31
Ventura, CA	3	163	164	0	-9	22
San Joaquin Valley	2	155	145	-7	-3	16
Sacramento, CA	1	143	139	-3	-12	18
Phoenix, AZ	2	149	133	-10	-1	22
Houston, TX	12	179	176	2	-22	35
Beaumont, TX	10	161	138	-11	-16	27
Dallas, TX	9	150	145	-3	-5	19
Baton Rouge, LA	7	137	150	9	-5	22
St. Louis, MO	8	147	143	-2	-1	21
Lake Michigan	8	150	145	-3	3	16
Detroit, MI	4	156	146	-7	-16	23
Nashville, TN	4	137	125	-9	-13	20
Louisville, KY	3	145	162	11	-7	21
Cincinnati, OH	6	152	166	10	-5	21
Atlanta, GA	4	172	183	7	0	20
Richmond, VA	6	151	137	-9	-7	19
Philadelphia, NJ	5	188	230	23	17	30
New York, NY	17	183	200	14	13	27
Baltimore, MD	4	162	165	3	0	19
New England	2	203	199	-2	-11	25
Average	5.9	164.4	164.2	0.3	-5.6	22.7
Standard Deviation	3.9	29.7	36.7	8.8	8.9	4.9
Minimum	1	137	125	-11	-22	16
Maximum	17	278	286	23	17	35

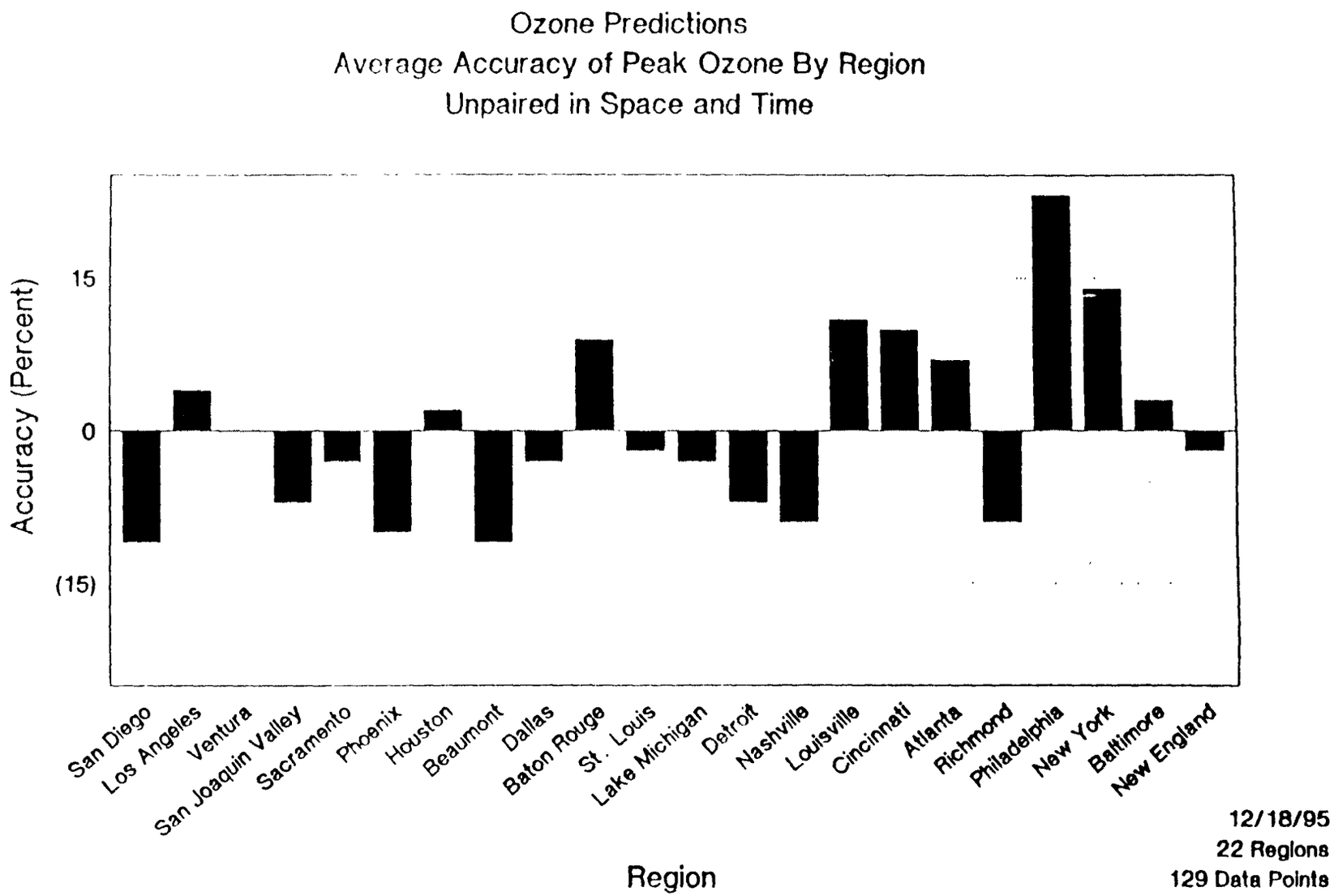

Figure 3-4. The average accuracy of the domain-wide peak ozone unpaired in space and time in each region.

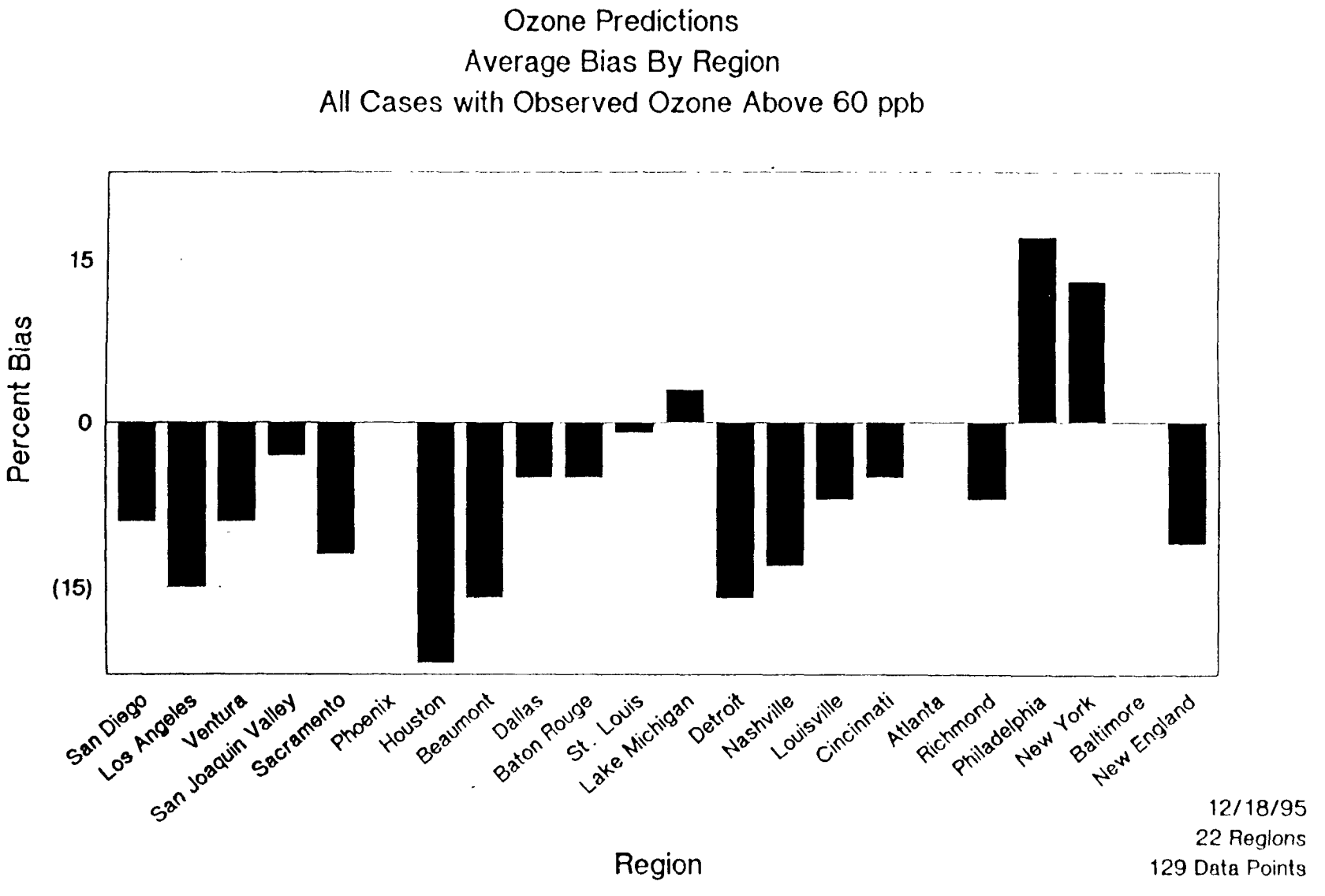

Figure 3-5. The average normalized bias in predicted ozone concentrations above 60 ppb in each region.

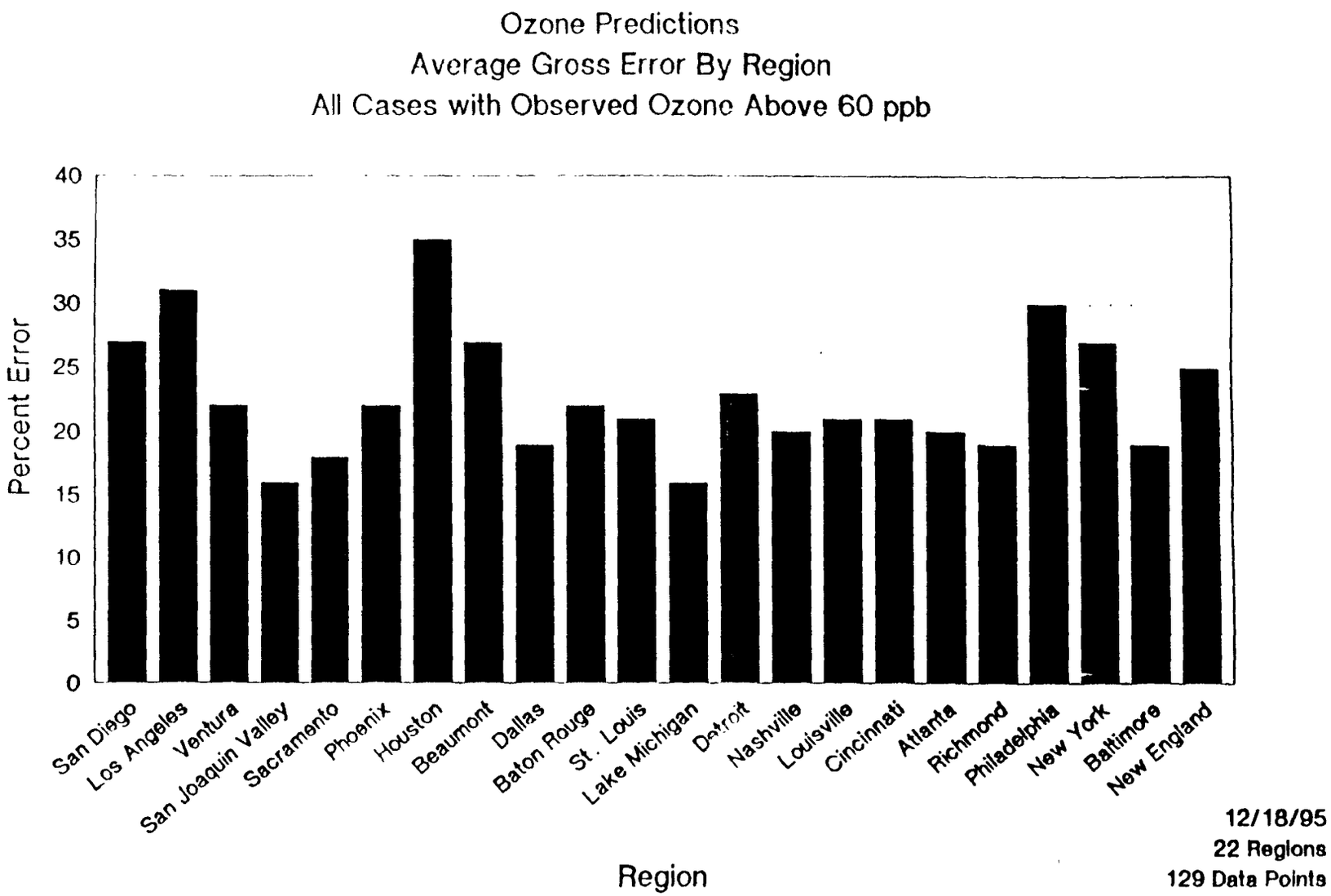

Figure 3-6. The average normalized gross error in predicted ozone concentrations above 60 ppb in each region.

normalized bias. Overall, 75 percent of the individual simulations met this performance goal. For

3.3 SIMULATION STATISTICS ON THE HIGHEST OZONE DAY IN EACH REGION

Table 3-4 and **Figures 3-7 through 3-9** show the model performance statistics for the highest ozone day in each region. The results show the domain-wide peak ozone was underpredicted by 6.6 percent on average on the highest ozone days in each region. On average, the predicted and observed domain-wide ozone peak concentrations were 183 and 196 ppb on the highest days in each region. The mean bias and gross error averaged -4.8 and 23.7 percent on the highest ozone days, respectively. The correlation between the predicted and observed peaks on the highest ozone days was moderately high, $R^2 = 0.70$. Thus, overall, the performance on the highest ozone days is not significantly different than the average of all days modeled. The only difference is the tendency for the models to underpredict the domain-wide peak on the highest days, which was especially evident in New York and Houston.

3.4 SIMULATION STATISTICS STRATIFIED BY MODELING METHODOLOGY AND GEOGRAPHIC SETTING

Within the modeling community, there is interest in whether the modeling methodologies and geographic setting significantly influence the model performance. There are many minor differences in the modeling methodologies in these simulations and in the geographic settings of the areas. Often these differences are difficult to quantify and classify consistently. For this analysis, we have elected to stratify only on a few variables.

The results of these simulations were stratified on (1) the type of wind model, prognostic or diagnostic, used to develop the hourly gridded 3-dimensional windfields and (2) the methods of determining the boundary conditions, either derived from a regional model (ROM) or from observations. In addition, the results were stratified on (1) whether the geophysical features were primarily flat land or involved a land-water interface, which generally increases the complexity of the meteorology, and (2) whether regional pollutant transport typically has a major or minor effect on ozone in the region. **Table 3-5** lists the boundary condition methodology, the wind model, the type of geographic setting, and the predominance of pollutant transport into the region. The data indicate that prognostic wind models were used more frequently in regions with more complex meteorology, so these stratification variables are not necessarily independent of one another. It is also important to recognize that the later classifications, especially the importance of pollutant transport into the region, are subjective and may not be accurate for all of the episodes in a region.

As shown in **Table 3-6** and **Figure 3-10**, the stratification on the boundary condition methodologies indicate a tendency towards underprediction when boundary conditions were derived from observations rather than from a regional model. The mean bias averaged -9 percent in the 78 simulations where boundary conditions were derived from observations compared to +1.5 percent mean bias on average in the 51 simulations with ROM-derived boundary conditions. The gross error was not significantly affected by the boundary condition methodology. The

Table 3-4. Photochemical models' ozone performance statistic on high ozone day by region.

Region	Observed Domain-wide Peak (ppb)	Predicted Domain-wide Peak (ppb)	Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias (%)	Mean Gross Error (%)
San Diego, CA	156	135	-13	-6	31
Los Angeles, CA	360	392	9	7	33
Los Angeles, CA	360	382	6	7	32
Ventura, CA	180	173	-4	-2	21
San Joaquin Valley	160	150	-6	-7	15
Sacramento, CA	143	139	-3	-12	18
Phoenix, AZ	157	134	-15	-5	24
Houston, TX	240	112	-53	-45	45
Beaumont, TX	226	151	-33	-32	32
Dallas, TX	170	155	-9	-6	23
Baton Rouge, LA	157	160	2	1	21
St. Louis, MO	163	160	-2	-7	20
Lake Michigan	178	150	-16	5	14
Detroit, MI	168	164	-2	-12	22
Nashville, TN	138	139	1	-9	20
Nashville, TN	138	115	-17	-7	15
Louisville, KY	151	187	24	2	16
Cincinnati, OH	169	169	0	4	19
Atlanta, GA	201	194	-3	5	28
Richmond, VA	173	146	-16	-16	19
Philadelphia, NJ	210	201	-4	7	26
Philadelphia, NJ	210	274	30	20	31
New York, NY	294	206	-30	-3	29
Baltimore, MD	180	169	-6	3	15
New England	222	211	-5	-12	24
Mean (weighted by region)	196.2	182.7	-6.6	-4.8	23.7
Standard Deviation	59.8	68.9	16.5	12.8	7.3
Minimum	138.0	112.0	-53.3	-45.0	14.0
Maximum	360.0	392.4	30.4	20.1	45.2

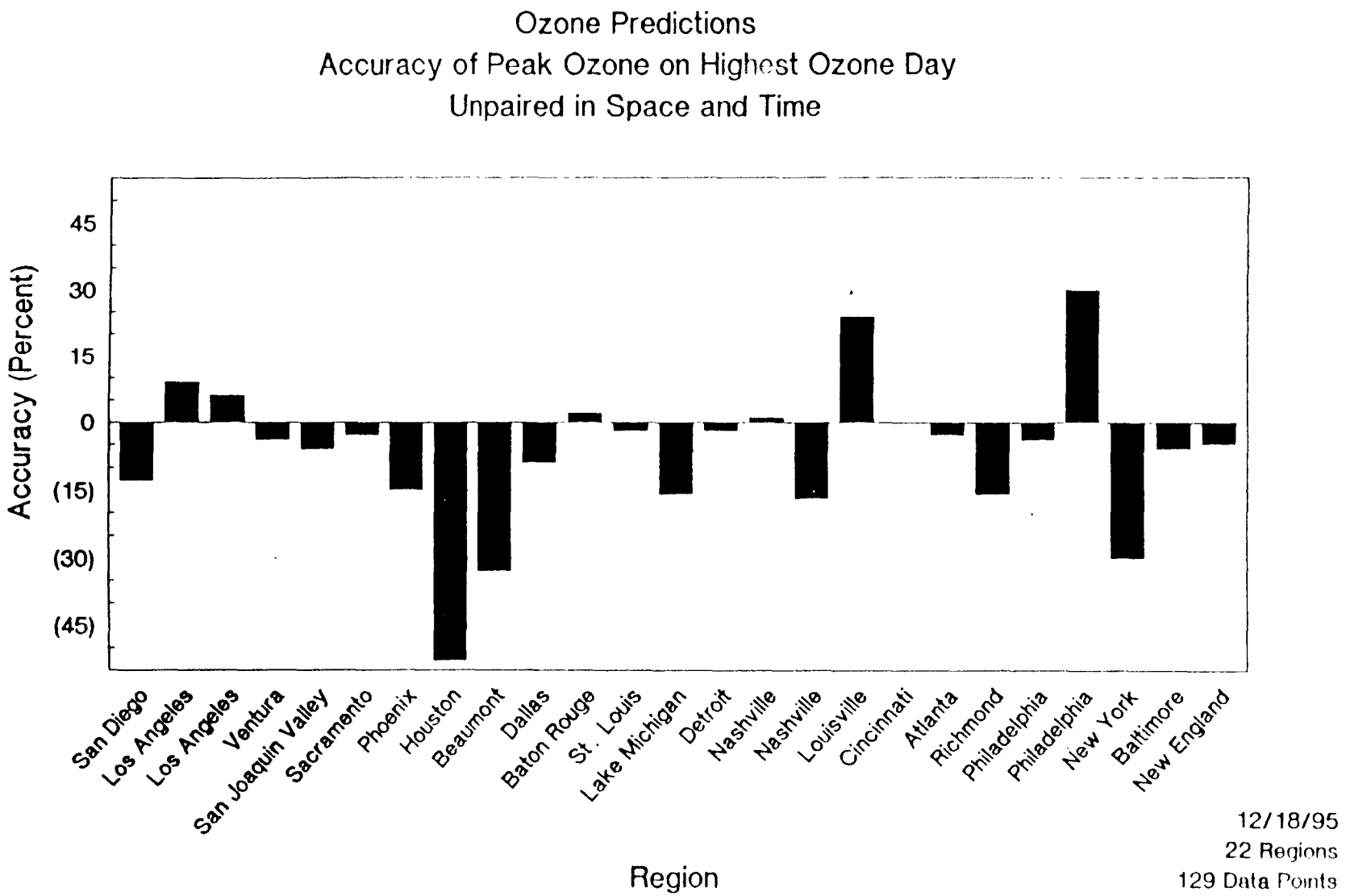

Figure 3-7. The accuracy of the domain-wide peak ozone unpaired in space and time on the highest ozone day in each region.

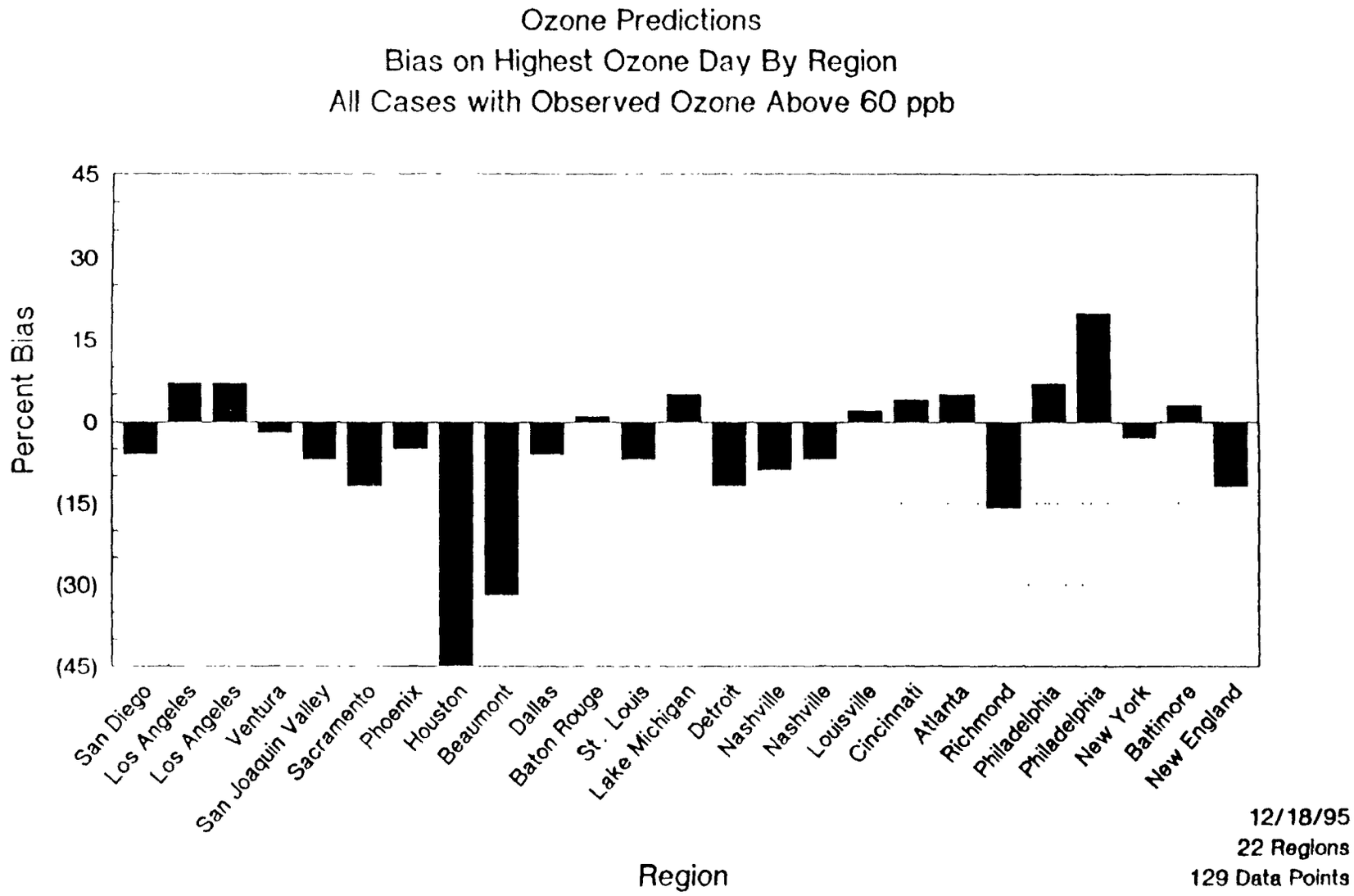

Figure 3-8. The mean normalized bias in predicted ozone concentrations above 60 ppb on the highest ozone day in each region.

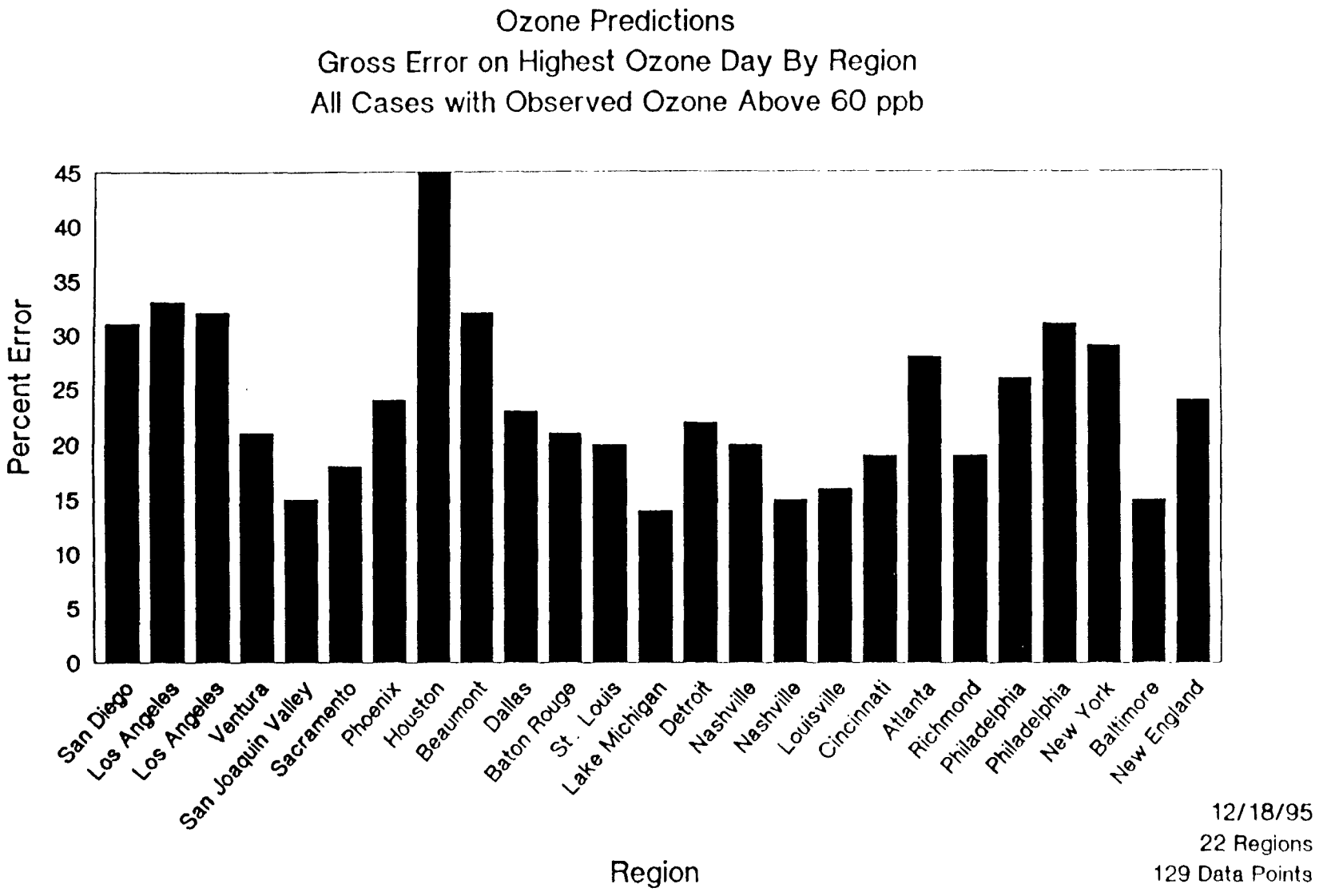

Figure 3-9. The mean normalized gross error in predicted ozone concentrations above 60 ppb on the highest ozone day in each region.

Table 3-5. The methodologies used to develop boundary conditions and windfields, and the characteristics of the geographic setting.

Region	Boundary Condition Method	Windfield Method	Geographic Setting	Inflow Pollutant Transport
San Diego, CA	Surface & Aircraft Observations	DWM	Land-water	Minor
Los Angeles, CA	Surface & Aircraft Observations	SAIMM-4DDA, DWM, CSUMM, HPEM	Land-water	Minor
Ventura, CA	Surface Observations	DWM	Land-water	Minor
San Joaquin Valley	Surface & Aircraft Observations	MM5	Land-water	Minor
Sacramento, CA	Surface & Aircraft Observations	DWM	Flat Land	Minor
Phoenix, AZ	Surface Observations	DWM	Flat Land	Minor
Houston, TX	Surface Observations	CSUMM	Land-water	Minor
Beaumont, TX	Surface Observations	CSUMM	Land-water	Minor
Dallas, TX	Surface Observations	DWM	Flat Land	Minor
Baton Rouge, LA	Surf & Aircraft	DWM	Land-water	Minor
St. Louis, MO	Surface Observations	DWM	Flat Land	Minor
Lake Michigan	Surface & Aircraft Observations	CALRAMS	Land-water	Major
Detroit, MI	ROM	CSUMM	Land-water	Major
Louisville, KY	ROM	DWM	Flat Land	Major
Nashville, TN	ROM	ROM	Flat Land	Major
Cincinnati, OH	ROM	DWM	Flat Land	Major
Atlanta, GA	Surface Observations	SAIMM	Flat Land	Minor
Richmond, VA	ROM	DWM	Land-water	Minor
Philadelphia, NJ	ROM	DWM	Land-water	Major
New York, NY	ROM	CALMET	Land-water	Major
Baltimore, MD	ROM	DWM	Land-water	Major
New England	ROM	CALMET	Land-water	Major

conditions and underpredicted by 1 percent with observation-based boundary conditions. Further stratification of the results based on whether only surface observations or surface and aircraft observations were used to determine boundary conditions showed only minor differences in the performance statistics. However, within the group of simulations with boundary conditions derived from observations, the model performance was slightly better when aircraft and surface observations were used instead of surface data alone.

Stratification of the model performance on the type of wind model employed showed larger differences, as shown in Table 3-6 and **Figure 3-11**. About one third of the simulations were made using winds derived from prognostic wind models. Simulations made with winds derived from prognostic models had more negative bias and larger error than those made with winds derived from diagnostic models. On average, the mean bias was -12 and -0.7 percent in simulations made with winds derived from prognostic and diagnostic models, respectively. The mean gross error was 26 and 23 percent in simulations made with winds derived from prognostic and diagnostic models, respectively. The differences in the average domain-wide peak accuracy were consistent with the other results but smaller (-2 percent with prognostic winds and +4.3 percent with diagnostic winds). It is also worth noting that ozone model performance with the UAM-DWM wind model was comparable to those for all diagnostic wind models. The under- and overpredictions of ozone on average may be due to the tendencies for prognostic and diagnostic models to overestimate and underestimate wind speeds, respectively. The poorer performance of simulations made with prognostic winds may also reflect the greater complexity of meteorology in the regions where the modelers elected to use this approach. Thus, the results should be interpreted cautiously because this stratification does not compare wind models in the same regions and, therefore, other confounding factors, such as the complexity of meteorology or emission inventory problems, may be responsible for the differences.

Stratification of the results on the basis of whether the geophysical characteristics include only land or land and water also shows differences, as shown in Table 3-6 and **Figure 3-12**. The average gross error in the predictions are 20 and 26 percent on average in regions with only land and with land and water, respectively. The other statistics are comparable. Not surprisingly, these results indicate that there is more model error (not bias) in regions with complex meteorology.

Lastly, stratification on the extent of pollutant transport into the region indicates there is negative bias in the regions where regional pollutant transport is a minor factor (see Table 3-6 and **Figure 3-13**). On average, the mean bias is -10 percent in these simulations compared to +3 percent in the simulations where regional pollutant transport is typically a more important factor. The domain-wide peak ozone predictions are more accurate on average in regions where pollutant transport is a minor factor. However, these differences are not large and should be interpreted cautiously because the stratification is subjective.

Table 3-6. Photochemical models' performance statistics for ozone stratified by boundary condition methodology, windfield model, geographic setting, and predominance of pollutant transport into the region.

Stratification Parameter	Number of Episode Days	Mean Domain-wide Peak Observation (ppb)	Mean Domain wide Peak Prediction (ppb)	Mean Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias (%)	Mean Gross Error (%)
Boundary conditions derived from the Regional Oxidant Model	51	167.3	176.3	6.4	1.5	23.7
Boundary conditions derived from observations	78	172.5	169.3	-1.0	-9.0	24.5
Boundary conditions derived from surface and aircraft observations	25	200.8	201.0	-0.3	-5.9	23.1
Boundary conditions derived from surface observations	54	159.2	154.3	-1.4	-10.4	25.0
Windfield derived from Prognostic Wind Model	48	185.1	179.5	-2.0	-11.7	26.3
Windfield derived from Diagnostic Wind Models	81	161.8	167.6	4.3	-0.7	22.9
Windfield derived from the UAM-DWM	60	157.4	161.3	2.5	-3.9	21.9
Land-only geographic setting	37	149.9	150.4	0.5	-4.7	20.2
Land-water geographic setting	92	178.7	180.8	2.5	-4.8	25.8
Significant pollutant transport into region	54	166.4	175.2	6.4	2.6	23.3
Minor pollutant transport into region	73	173.4	169.8	-1.2	-10.2	24.8
Average	129	170.5	172.1	1.9	-4.8	24.2

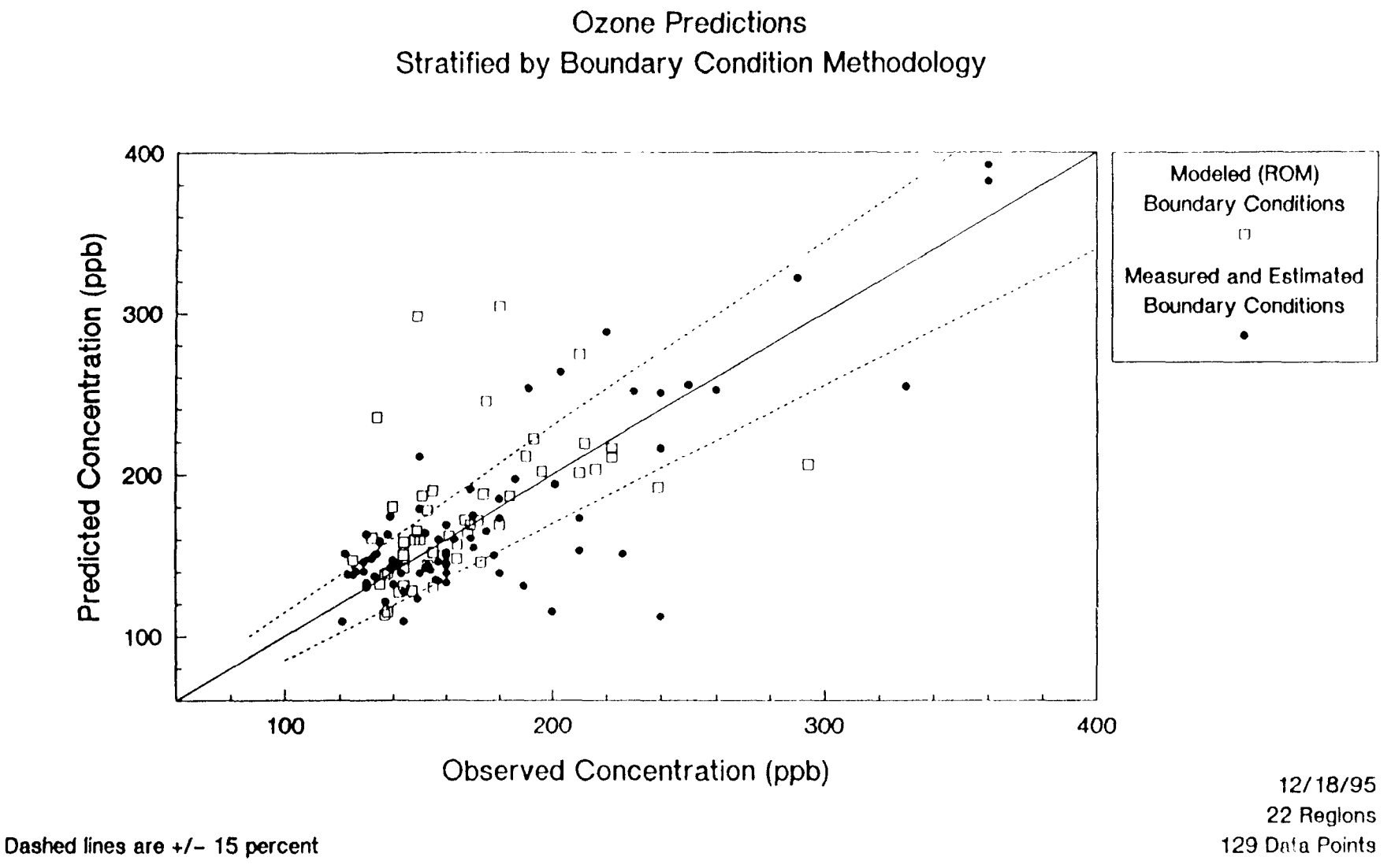

Figure 3-10. The predicted and observed domain-wide maximum ozone concentrations stratified by boundary condition methodology.

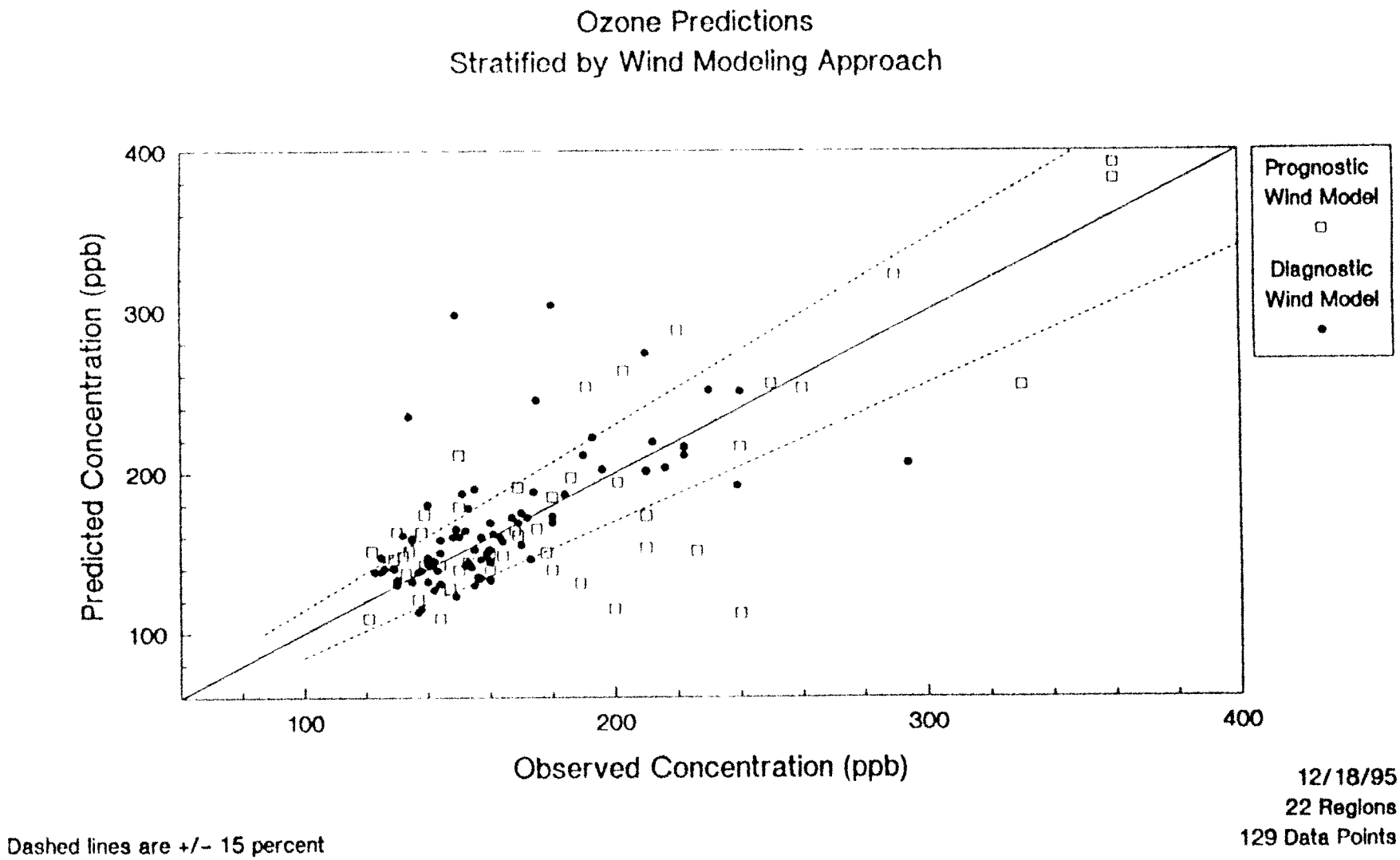

Figure 3-11. The predicted and observed domain-wide maximum ozone concentrations stratified by wind model type.

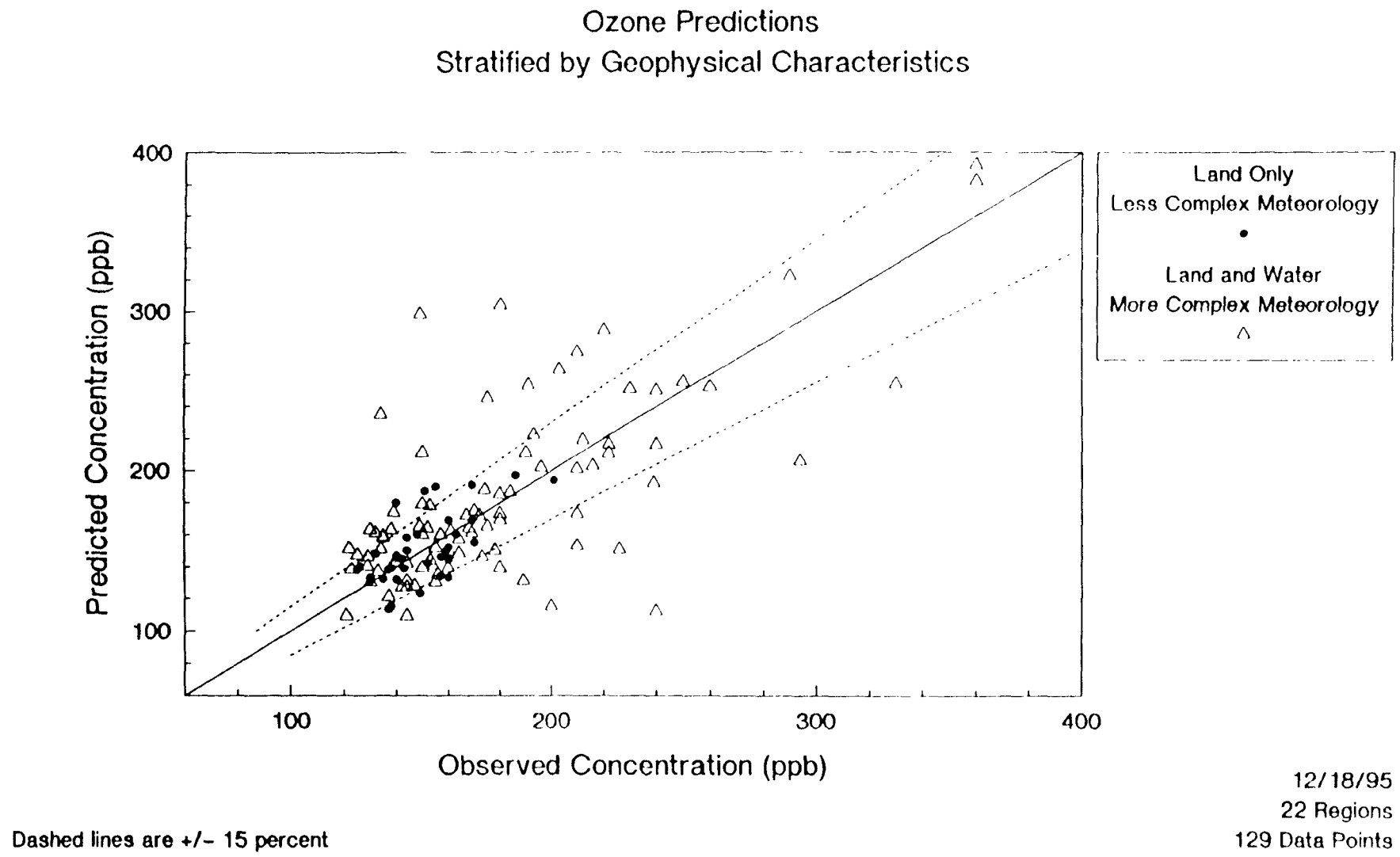

Figure 3-12. The predicted and observed domain-wide maximum ozone concentrations stratified by geophysical characteristics.

3.5 SIMULATION STATISTICS FOR THE UAM-IV, UAM-V, AND SAQM MODELS

As indicated above, most of the simulations were made with the UAM-IV model and there was only one region, Atlanta, where a direct comparison of the UAM-IV and UAM-V models was made using the same input database (SAI, 1994c). **Table 3-7** shows the average ozone performance statistics for four days of simulations in Atlanta with the two models. The results show the mean normalized bias and gross error are virtually identical in these simulations (0.0 vs. -0.6 percent bias and 20.4 vs. 20.9 percent gross error). However, the average accuracy of the domain-wide peak ozone from UAM-IV and UAM-V models is different, but still reasonably accurate: +6.9 percent with the UAM-IV model and -3.9 percent with the UAM-V model. Overall, these results indicate the performance of both models is excellent and comparable in the Atlanta simulations. These results should be interpreted cautiously because the characteristics of the Atlanta area and ozone episodes probably do not represent a stressful test of the two models. That is, technical improvements incorporated into UAM-V may affect the performance more significantly in areas with more complex meteorology or emission patterns, such as Los Angeles or Lake Michigan. The good agreement found in the Atlanta simulations may not hold for other areas.

Table 3-7 Comparison of the average model performance statistics for ozone from the SAQM, UAM-IV, and UAM-V models

Model/Region	Number of Days	Mean Accuracy of Domain-wide Peak Unpaired (%)	Mean Normalized Bias (%)	Mean Gross Error (%)
UAM-IV in Atlanta	4	6.9	0	20.4
UAM-V in Atlanta	4	-3.9	-0.6	20.9
SAQM in San Joaquin Valley	2	-6.8	-3.0	15.5
UAM-V in Lake Michigan	8	-3.1	3.4	16.1
UAM-IV in 20 Areas	119	2.4	-5.4	24.9

Table 3-7 also shows a comparison of the average performance statistics for the SAQM model in the San Joaquin Valley simulations and the UAM-IV model in the 20 other regions. The statistics show the SAQM simulations have significantly less gross error than the UAM-IV simulations, but the domain-wide peak ozone is also less accurate than the average from all of the UAM-IV simulations. The low gross error and near absence of bias in the SAQM simulations is impressive performance and reflects not only the strengths of the SAQM/MM5 models but also the extensive aerometric database available for the episodes and the extensive effort to refine model performance. A process-oriented approach including multi-species comparisons was used

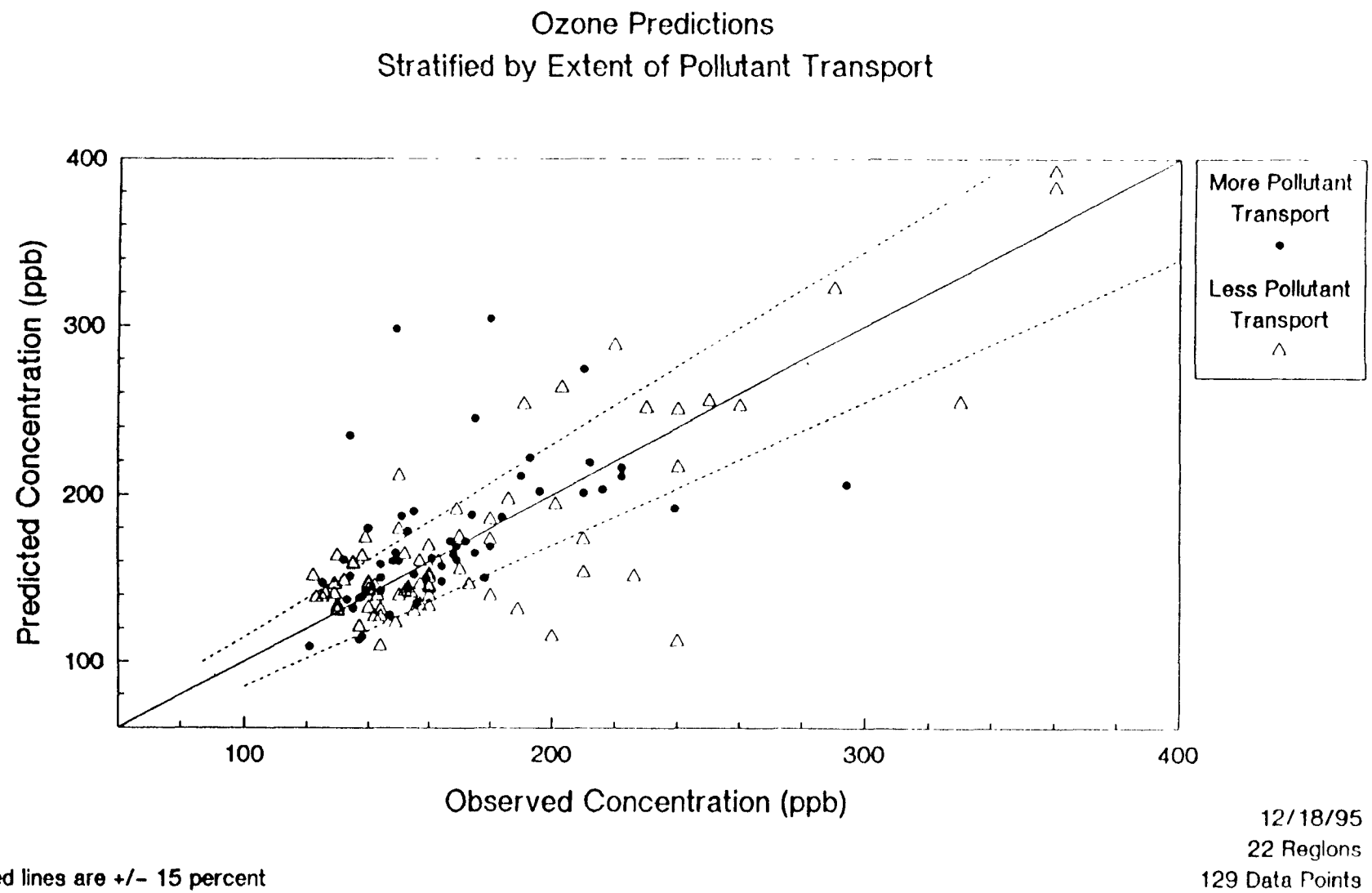

Figure 3-13. The predicted and observed domain-wide maximum ozone concentrations stratified by the predominance of pollutant transport into the region.

over a two-year period to improve model performance for the SAQM simulations and the performance statistics reflect the large effort.

Average performance statistics for the UAM-V model in the Lake Michigan simulations are also shown in Table 3-7. The accuracy of the domain-wide peak predictions and the mean normalized bias are comparable to those for the UAM-IV in the 20 regions; however, the gross error in the UAM-IV Lake Michigan simulations is significantly lower than the average for the UAM-IV simulations in other areas (16.1 vs 24.9 percent). Like in the San Joaquin Valley, an extensive special study aerometric database was available for these episodes and a process-oriented approach including multi-species comparisons was used over a two-year period to improve model performance for the Lake Michigan simulations. The UAM-V ozone performance is clearly better than the average UAM-IV performance in other areas on average, and the difference is probably due to both features of the UAM-V model and the better database and more extensive model evaluation procedures employed in the Lake Michigan modeling.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 CONCLUSIONS

The investigation of photochemical models ozone performance in the SIP applications is based on three basic performance statistics required by EPA for regulatory applications. The compilation results for 131 simulations in 22 regions showed the average accuracy of the domain-wide peak ozone unpaired in space and time was +2 ±20 percent. The mean normalized bias in concentrations above 60 ppb in all simulations was - 5 ±16 percent. The mean normalized gross error in concentrations above 60 ppb in all simulations was 24 ±8 percent.

The EPA established model performance goals for regulatory applications of the UAM. The goals were for the unpaired peak accuracy and mean normalized bias to be within ±15 percent, and for the mean gross error to be less than 35 percent. Almost all (94 percent) of the simulations met the gross error goal and over half (64 percent) of the simulations met all three goals. Furthermore, at least one simulation in each region except Houston met all three goals. Thus, while there were areas with better and poorer model performance, almost all areas had at least one episode day with acceptable performance (based on the EPA goals) which could be used for control strategy evaluations.

Stratification of the results on selected methodological procedures did not show large differences. Simulations made with boundary conditions derived from observations predicted lower ozone on average than those with boundary conditions derived from the Regional Oxidant Model. Simulations made with prognostic wind models had larger error and more negative bias on average compared to simulations made with windfields derived from diagnostic wind models, however, the areas where prognostic models were employed are also regions with more complex meteorology and where larger model error would be expected. The negative bias in ozone with the prognostic models may be due to their tendency to overestimate wind speeds, but it also could be due to other factors.

Stratification of the ozone performance statistics on geographic characteristics also showed modest differences. For example, there was greater model error on average in regions with land-water interfaces (26 percent versus 20 percent). This result was expected because regions with land-water interfaces have more complex meteorology. However, regions where pollutant transport into the region was expected to be significant had lower model error and less bias than those without significant pollutant transport. This result may suggest inflow boundary conditions were adequately estimated in the regions with significant transport in these simulations, but there could also be other compensating factors which explain the differences. Thus, all of the stratification results should be interpreted cautiously.

A comparison of results from applications of the UAM-IV and UAM-V models to the Atlanta regions showed excellent and comparable performance. The average results for the four simulated days differed only in the accuracy of the domain-wide peak predictions, which were slightly better with the UAM-V model. Overall, the results were too similar to distinguish

between the models. The characteristics of the Atlanta region and episodes may not be sufficient to stressfully test the differences between the UAM-IV and UAM-V models, so the absence of differences in these test should be interpreted cautiously.

A comparison of the SAQM model performance in the San Joaquin Valley and the UAM-V model performance in Lake Michigan to the average UAM-IV model performance in 20 other regions showed the SAQM and UAM-V models had significantly less gross error in predicting ozone concentrations above 60 ppb. The accuracy of the domain-wide peak concentrations and the mean normalized bias were similar for the SAQM, UAM-V, and UAM-IV on average. The better overall performance of the SAQM and UAM-V models in these simulations is probably due to both features of the models and the better databases and more extensive model evaluation procedures employed in the modeling in these two areas. Process-oriented approaches including multi-species comparisons were used to improve model performance for the simulations and the performance statistics reflect the more comprehensive efforts.

4.2 RECOMMENDATIONS

The analysis of photochemical grid model performance using three statistics for ozone provides a good starting point for model evaluation, but it does not represent a comprehensive approach to model evaluation A more thorough approach would include the following elements

- Use of additional model performance statistics, such as those listed in **Table 4-1**

- Evaluation of the model performance for precursor species, such as NO, NO_2, NO_y, CO, lumped VOC species (OLE, PAR, TOL, XYL, ETH, FORM, and ALD), in addition to those for ozone

- Use of a process-oriented approach that reviews all of the statistical results and graphical displays for baseline runs, alternate baseline runs, and other sensitivity runs, and seeks to identify potential problems or inconsistencies, including compensating errors.

In order to carry out more comprehensive analyses, the evaluation team must have access to all of the model input, model output, and observation files used in the simulations and all of the modeling documentation, and have appropriate display software to visualize the predictions and observations

Table 4-1 Additional statistical parameters for model performance evaluation.

Abbreviation	Parameter
PAPST	Absolute accuracy of domain-wide maximum 1-hr concentrations paired in space and time*
NPAPST	Normalized accuracy of domain-wide maximum 1-hr concentrations paired in space and time*
PAPS	Absolute accuracy of domain-wide maximum 1-hr concentrations paired in space and unpaired in time*
NPAPS	Normalized accuracy of domain-wide maximum 1-hr concentrations paired in space and unpaired time*
PAPT	Absolute accuracy of domain-wide maximum 1-hr concentrations paired in time and unpaired in space*
NPAPT	Normalized accuracy of domain-wide maximum concentrations paired in time and unpaired in space*
PA	Absolute accuracy of domain-wide maximum 1-hr concentrations unpaired in space and time*
PNBIAS	Mean normalized bias of predicted and observed maximum 1-hr concentrations at all monitoring stations*
PNERROR	Mean normalized error of predicted and observed maximum 1-hr concentrations at all monitoring stations*
PBIAS	Mean absolute bias of predicted and observed maximum 1-hr concentrations at all monitoring stations*
PERROR	Mean absolute error of predicted and observed maximum 1-hr concentrations at all monitoring stations*
BIAS	Mean absolute bias of all predicted and observed concentration pairs where the observed exceeds a minimum concentration
ERROR	Mean absolute error of all predicted and observed concentration pairs where the observed exceeds a minimum concentration

* These definitions assume all observed peak concentrations exceed the minimum concentrations

5. REFERENCES

CARB (1995a) Revisions to the base case and future year Urban Airshed Model simulations for Ventura County in support of the 1994 State Implementation Plan. Report prepared by California Air Resources Board, Sacramento, CA.

CARB (1995b) Photochemical modeling of the Greater Sacramento Area in support of the 1994 State Implementation Plan. Report prepared by California Air Resources Board, Sacramento, CA.

CARB (1995c) San Joaquin Valley SIP Modeling: Model Performance. Report prepared by California Air Resources Board, Sacramento, CA.

EPA (1991) Guidelines for regulatory application of the Urban Airshed Model. Report prepared by U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA450/4-91-013.

Georgopoulos P G (1995) Ozone sip modeling technical support documentation summary for the New Jersey - Philadelphia CMSA area. Report prepared by Ozone Research Center, Environmental and Occupational Health Sciences Institute, Piscataway, NJ, Technical ORC-TR9502.

Kaminski M.A. (1995) Urban airshed modeling of the middle Tennessee modeling domain: model performance evaluation. Report prepared by University of Tennessee, Knoxville, TN

LADCO (1994) Lake Michigan Ozone Study: evaluation of the UASM-V photochemical grid model in the Lake Michigan region, Version 2.0. Report prepared by Lake Michigan Air Directors Consortium, Des Plaines, IL.

NYDEC (1994a) New York urban airshed modeling For June 14 to 20, 1987 base case. New York State Division of Environmental Conservation, Albany, NY, NYAS-94-SE3.

NYDEC (1994b) New York urban airshed modeling For July 5 to 11, 1988 base case. New York State Division of Environmental Conservation, Albany, NY, NYAS-94-SE1.

NYDEC (1994c) New York urban airshed modeling For June 18 to 22, 1988 base case. New York State Division of Environmental Conservation, Albany, NY, NYAS-94-SE2.

NYDEC (1994d) New York urban airshed modeling For July 16 to 20, 1991 base case. New York State Division of Environmental Conservation, Albany, NY, NYAS-94-SE4.

Roth P.M., Reynolds S., Tesche T., and Dennis R. (1991) A conceptual framework for evaluating the performance of grid-based photochemical air quality simulation models. Report prepared by Envair, San Anselmo, CA.

SAI (1990) User's guide to the Urban Airshed Model (UAM-IV). Report prepared by Systems Applications International, San Rafael, CA.

SAI (1993) User's guide: Lake Michigan Ozone Study photochemical modeling system. Report prepared by Systems Applications International, San Rafael, CA.

SAI (1994a) Photochemical modeling of the Lake Michigan region for the 1991 Lake Michigan Ozone Study (LMOS) using the Nested-grid Urban Airshed Model (UAM-V). Report prepared by Systems Applications International, San Rafael, CA.

SAI (1994b) Application of the Urban Airshed Model to Baton Rouge, Louisiana for three multi-day ozone episodes. Volume III: diagnostic/sensitivity analysis and model performance evaluation. Report prepared by Systems Applications International, San Rafael, CA. SYSAPP94/090.

SAI (1994c) Comparison of the UAM-IV and UAM-V photochemical models for three Atlanta-area ozone episodes. Report prepared by Systems Applications International, San Rafael, CA. SYSAPP94/106.

SCAQMD (1994) 1994 Air Quality Management Plan, Technical Report V-B: Ozone Modeling Performance Evaluation. South Coast Air Quality Management District, Diamond Bar, CA

SDAPCD (1994) Request for ozone reclassification - urban airshed modeling technical support document. Report prepared by San Diego Air Pollutant Control District, San Diego, CA

Tesche T.W., Lurmann F.W., Roth P.M., Georgopoulos P., Seinfeld J.H., and Cass G. (1990) Improvements of procedures for evaluating photochemical models. Report prepared for California Air Resources Board, Radian Corp., Sacramento, CA, ARB Contract No. A832-103.

TNRCC (1994a) El Paso, Texas ozone nonattainment areas base case report performance evaluation. Report prepared by Texas Natural Resource Conservation Commission, Austin, TX.

TNRCC (1994b) Houston/Galveston Beaumont/Port Arthur ozone nonattainment areas base case report performance evaluation. Report prepared by Texas Natural Resource Conservation Commission, Austin, TX.

TNRCC (1994c) Dallas/Ft. Worth ozone nonattainment areas base case report performance evaluation. Report prepared by Texas Natural Resource Conservation Commission, Austin, TX.

TECHNICAL REPORT DATA
(Please read Instructions on reverse before completing)

1 REPORT NO EPA-454/R-96-004	2	3 RECIPIENT'S ACCESSION NO
4 TITLE AND SUBTITLE Compilation of Photochemical Models' Performance Statistics for 11/94 Ozone SIP Applications		5 REPORT DATE July 1996 6 PERFORMING ORGANIZATION CODE
7 AUTHOR(S)		8 PERFORMING ORGANIZATION REPORT NO
9 PERFORMING ORGANIZATION NAME AND ADDRESS Sonoma Technology Inc. 5510 Skylane Boulevard, Suite 101 Santa Rosa, CA 95403		10 PROGRAM ELEMENT NO 11 CONTRACT/GRANT NO EPA Contract No. 68D30020
12 SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency Office of Air Quality Planning and Standards Emissions, Monitoring & Analysis Division Research Triangle Park, NC 27711		13 TYPE OF REPORT AND PERIOD COVERED Final Report 14 SPONSORING AGENCY CODE

15 SUPPLEMENTARY NOTES
EPA Work Assignment Manager: Shao-Hang Chu

16 ABSTRACT

This report is a compilation of the model performance statistics of three photochemical models (UAM-IV, UAM-V, and SAQM) used in the 11/94 ozone State Implementation Plans (SIP) applications. The models were applied in 24 ozone nonattainment regions in 1993-1995 to support the development of emissions control strategies for the 1994 ozone SIPs. The evaluation focuses on three EPA recommended basic model performance statistical measures of the models' ability to predict ambient ozone concentrations. It does not include an evaluation of the models' performance on other species (such as NO, NO_2, NO_y, and VOCs) or other statistical measures on ozone. For a more complete evaluation of the models' performance both spatial and temporal analyses of the matching between the predicted and observed concentrations are desirable.

17 KEY WORDS AND DOCUMENT ANALYSIS

a DESCRIPTORS	b IDENTIFIERS/OPEN ENDED TERMS	c COSATI Field/Group
Ozone SIP Applications Modeling Photochemical Model Performance Statistics		
18 DISTRIBUTION STATEMENT Release Unlimited	19 SECURITY CLASS (Report) Unclassified 20 SECURITY CLASS (Page) Unclassified	21 NO OF PAGES 156 22 PRICE

EPA Form 2220-1 (Rev 4-77) PREVIOUS EDITION IS OBSOLETE