Ozone Forecast Tool Performance: June to August 2010

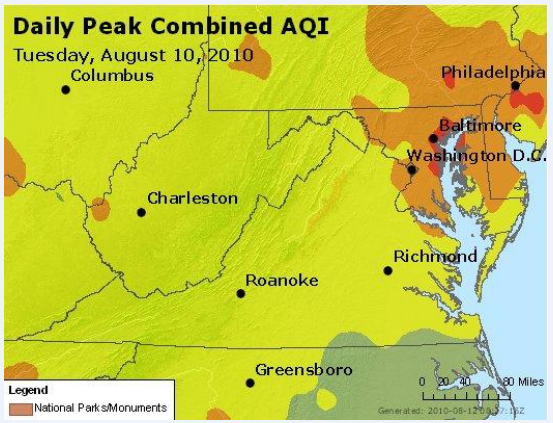
Debra Baker, MDE

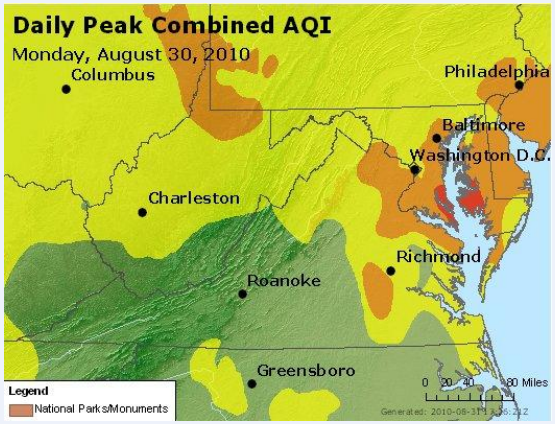
AQ Forecaster Focus Group Workshop

Silver Spring, MD September 16, 2010

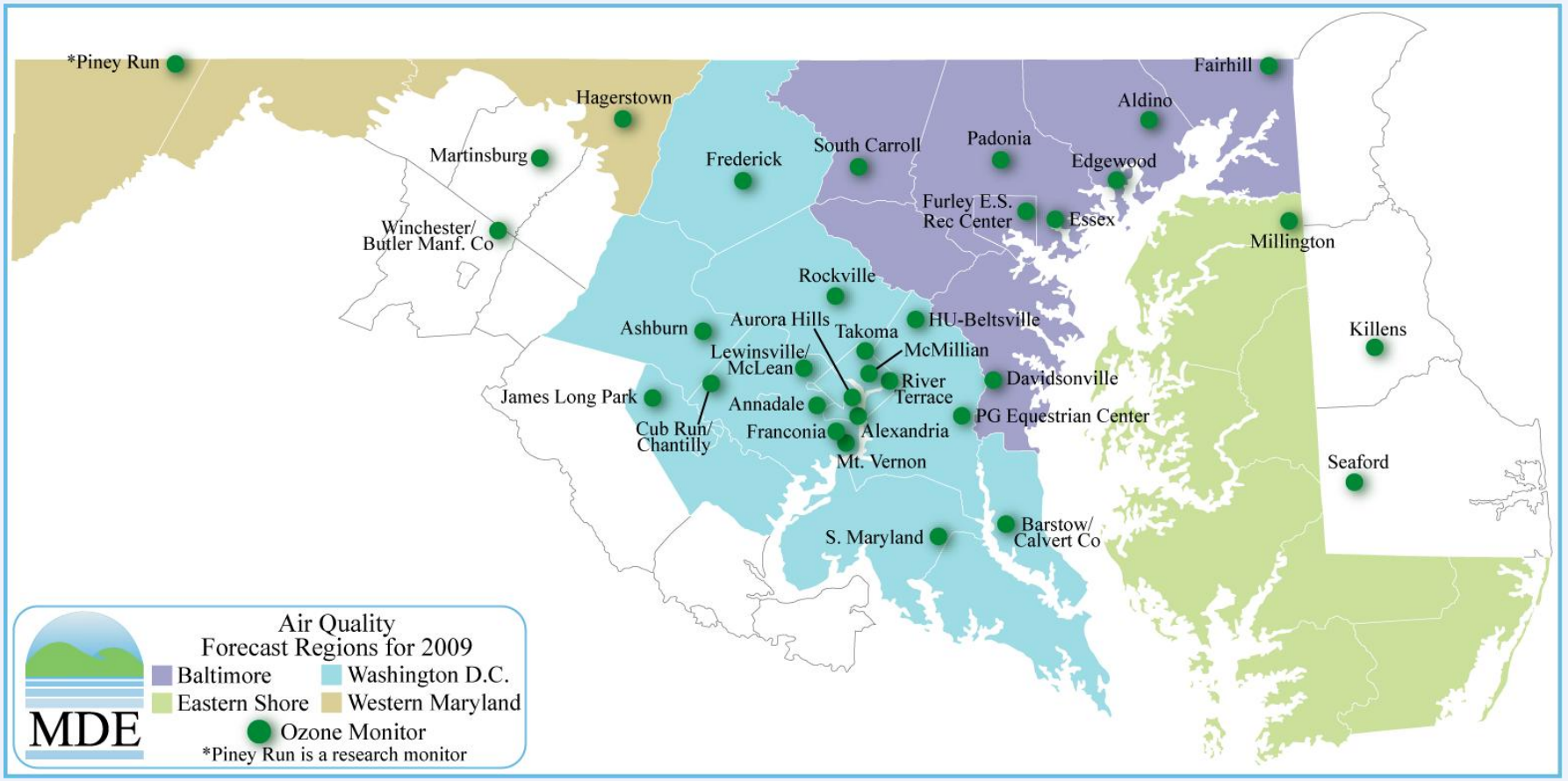
How useful are models for AQ forecasts?

- MDE has three air quality (AQ) forecasters who issue daily 3-day forecasts from April 15 to September 30 for both ozone and fine particulates ($PM_{2.5}$).
- Forecasters use several forecast tools, including numerical and statistical models. AQ-MOS is a new tool this year
- This study evaluates the accuracy of the ozone forecast tools for June, July, and August 2010.

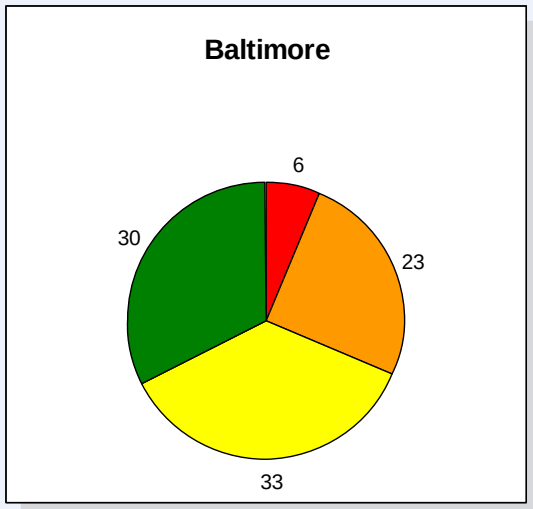


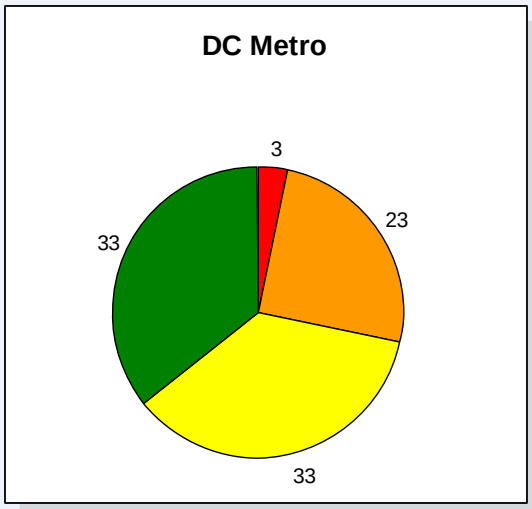


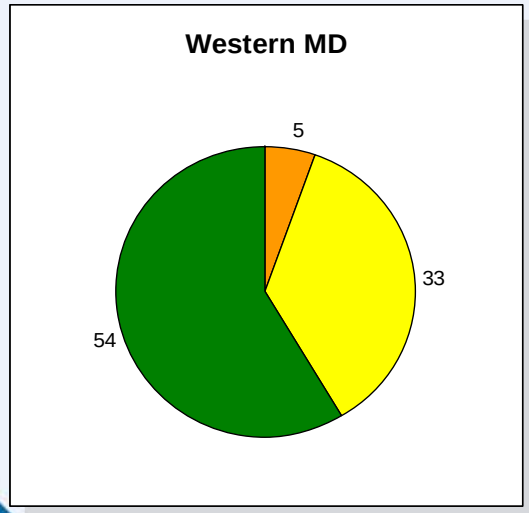
MDE's Four Forecast Regions

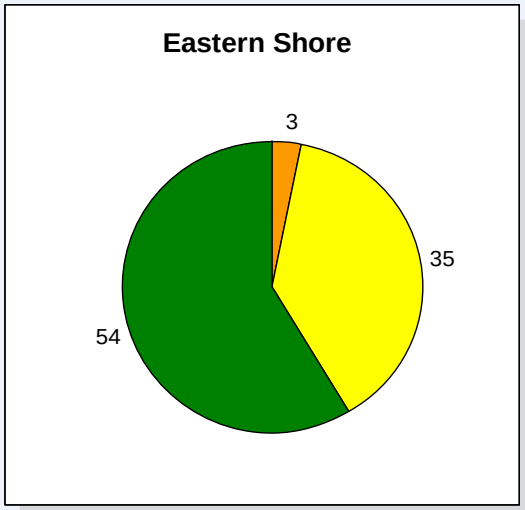


Ozone AQI Days for Summer 2010





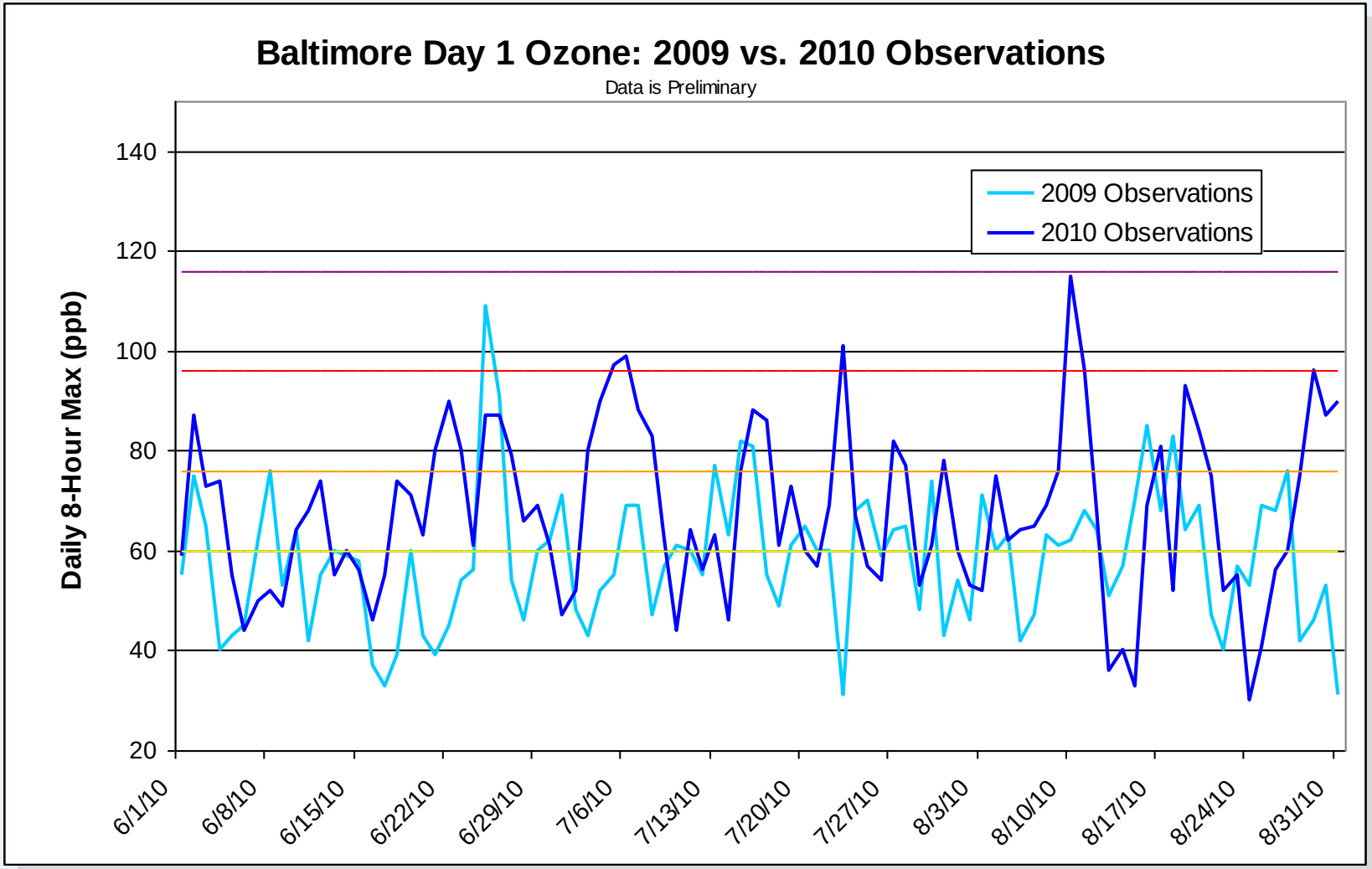




Average max 8-hour ozone 2009 → 2010

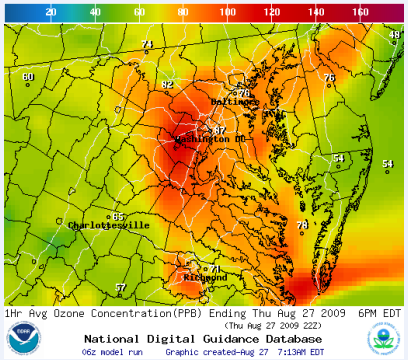
- BA 58 → 68 ppb
- DC 55 → 66 ppb
- WM 48 → 58 ppb
- ES 47 → 56 ppb

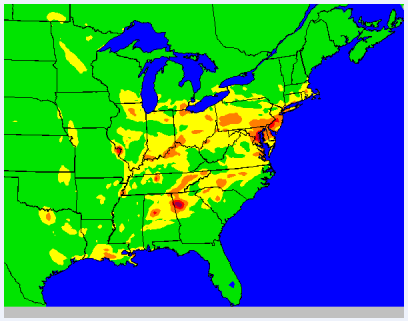
Major O_3 Increase 2009 to 2010

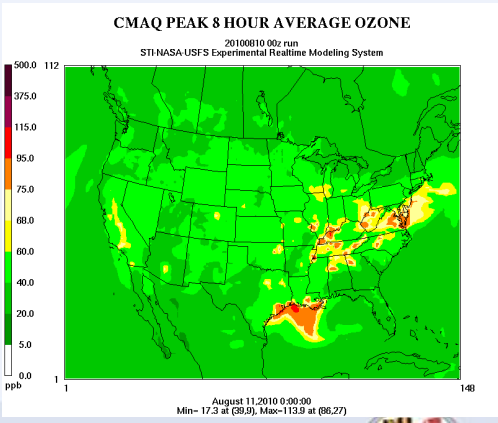


Ozone Forecast Tools

- Regional Regression Models (Ryan, PSU)
  - BA: R1 & R2
  - DC: R1, R2, & RB (average of R1 & R2)
  - WM: R1, R2, & RB (average of R2 & NOAA)
  - ES: R1 & R2
- NOAA Ozone Model
  - Forecast
  - AQ-MOS
- Blue Sky Ozone Model
  - Forecast
  - AQ-MOS
- Poor Man's Ensemble (Average of All)



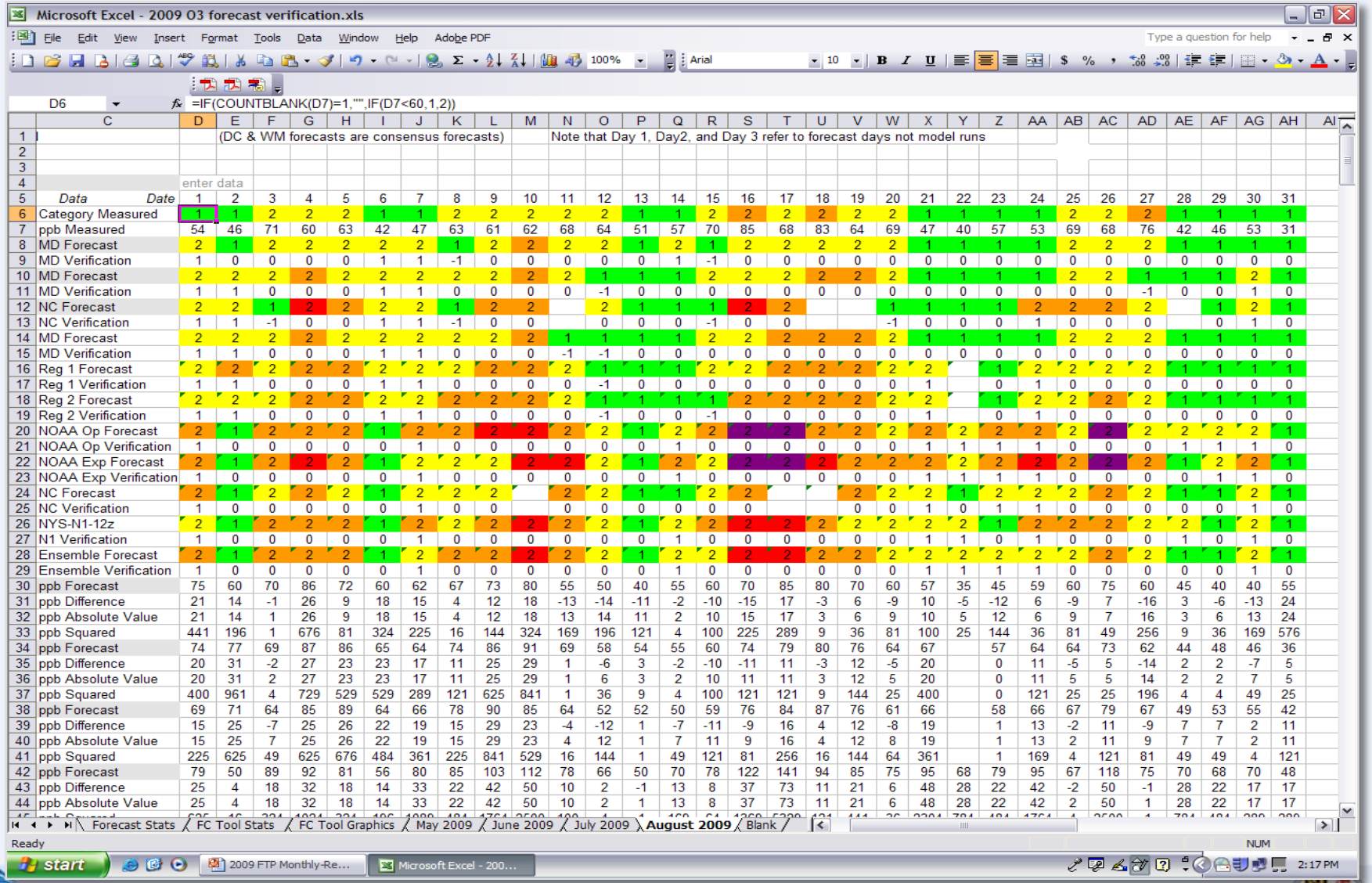




Methodology

- For NOAA and Blue Sky forecasts and AQ-MOS for each, point forecasts were used to determine the 8-hour maximum ozone for each region.
- The Regression Models also provided specific ozone concentrations for each region.
- The overall results presented here are an average for the four Maryland forecast regions.
- The highest daily monitor in each region was used as the observation for statistical verification.

Sample Spreadsheet



Microsoft Excel - 2009 O3 forecast verification.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF

D6 =IF(COUNTBLANK(D7)=1,"",IF(D7<60,1,2))

(DC & WM forecasts are consensus forecasts) Note that Day 1, Day2, and Day 3 refer to forecast days not model runs

enter data

Data	Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Category Measured		1	1	2	2	2	1	1	2	2	2	2	2	1	1	2	2	2	2	2	2	1	1	1	1	2	2	2	1	1	1	1
ppb Measured		54	46	71	60	63	42	47	63	61	62	68	64	51	57	70	85	68	83	64	69	47	40	57	53	69	68	76	42	46	53	31
MD Forecast		2	1	2	2	2	2	2	1	2	2	2	2	1	2	1	2	2	2	2	2	1	1	1	1	2	2	2	1	1	1	1
MD Verification		1	0	0	0	0	1	1	-1	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MD Forecast		2	2	2	2	2	2	2	2	2	2	2	1	1	1	2	2	2	2	2	2	1	1	1	1	2	2	1	1	1	2	1
MD Verification		1	1	0	0	0	1	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	1	0
NC Forecast		2	2	1	2	2	2	2	1	2	2		2	1	1	1	2	2			1	1	1	1	2	2	2	2		1	2	1
NC Verification		1	1	-1	0	0	1	1	-1	0	0		0	0	0	-1	0	0			-1	0	0	0	1	0	0	0		0	1	0
MD Forecast		2	2	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	2	1	1	1	1	2	2	2	1	1	1	1
MD Verification		1	1	0	0	0	1	1	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Reg 1 Forecast		2	2	2	2	2	2	2	2	2	2	2	1	1	1	2	2	2	2	2	2	2		1	2	2	2	2	1	1	1	1
Reg 1 Verification		1	1	0	0	0	1	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	1		0	1	0	0	0	0	0	0	0
Reg 2 Forecast		2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	2		1	2	2	2	2	1	1	1	1
Reg 2 Verification		1	1	0	0	0	1	1	0	0	0	0	-1	0	0	-1	0	0	0	0	0	1		0	1	0	0	0	0	0	0	0
NOAA Op Forecast		2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1
NOAA Op Verification		1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0
NOAA Exp Forecast		2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1
NOAA Exp Verification		1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	0
NC Forecast		2	1	2	2	2	1	2	2	2		2	2	1	1	2	2			2	2	2	1	2	2	2	2	2	1	1	2	1
NC Verification		1	0	0	0	0	0	1	0	0		0	0	0	0	0	0			0	0	1	0	1	1	0	0	0	0	0	1	0
NYS-N1-12z		2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1
N1 Verification		1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	1	0
Ensemble Forecast		2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1
Ensemble Verification		1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	0
ppb Forecast		75	60	70	86	72	60	62	67	73	80	55	50	40	55	60	70	85	80	70	60	57	35	45	59	60	75	60	45	40	40	55
ppb Difference		21	14	-1	26	9	18	15	4	12	18	-13	-14	-11	-2	-10	-15	17	-3	6	-9	10	-5	-12	6	-9	7	-16	3	-6	-13	24
ppb Absolute Value		21	14	1	26	9	18	15	4	12	18	13	14	11	2	10	15	17	3	6	9	10	5	12	6	9	7	16	3	6	13	24
ppb Squared		441	196	1	676	81	324	225	16	144	324	169	196	121	4	100	225	289	9	36	81	100	25	144	36	81	49	256	9	36	169	576
ppb Forecast		74	77	69	87	86	65	64	74	86	91	69	58	54	55	60	74	79	80	76	64	67		57	64	64	73	62	44	48	46	36
ppb Difference		20	31	-2	27	23	23	17	11	25	29	1	-6	3	-2	-10	-11	11	-3	12	-5	20		0	11	-5	5	-14	2	2	-7	5
ppb Absolute Value		20	31	2	27	23	23	17	11	25	29	1	6	3	2	10	11	11	3	12	5	20		0	11	5	5	14	2	2	7	5
ppb Squared		400	961	4	729	529	529	289	121	625	841	1	36	9	4	100	121	121	9	144	25	400		0	121	25	25	196	4	4	49	25
ppb Forecast		69	71	64	85	89	64	66	78	90	85	64	52	52	50	59	76	84	87	76	61	66		58	66	67	79	67	49	53	55	42
ppb Difference		15	25	-7	25	26	22	19	15	29	23	-4	-12	1	-7	-11	-9	16	4	12	-8	19		1	13	-2	11	-9	7	7	2	11
ppb Absolute Value		15	25	7	25	26	22	19	15	29	23	4	12	1	7	11	9	16	4	12	8	19		1	13	2	11	9	7	7	2	11
ppb Squared		225	625	49	625	676	484	361	225	841	529	16	144	1	49	121	81	256	16	144	64	361		1	169	4	121	81	49	49	4	121
ppb Forecast		79	50	89	92	81	56	80	85	103	112	78	66	50	70	78	122	141	94	85	75	95	68	79	95	67	118	75	70	68	70	48
ppb Difference		25	4	18	32	18	14	33	22	42	50	10	2	-1	13	8	37	73	11	21	6	48	28	22	42	-2	50	-1	28	22	17	17
ppb Absolute Value		25	4	18	32	18	14	33	22	42	50	10	2	1	13	8	37	73	11	21	6	48	28	22	42	2	50	1	28	22	17	17

Forecast Stats | FC Tool Stats | FC Tool Graphics | May 2009 | June 2009 | July 2009 | August 2009 | Blank

Ready NUM

start | 2009 FTP Monthly-Re... | Microsoft Excel - 200... | 2:17 PM

MDE

MARYLAND DEPARTMENT OF THE ENVIRONMENT

Categorical Tests

- Percentage Correct: (a+d)/(a+b+c+d)
- Critical Success Index: a/(a+b+c)
- Hit Rate: a/(a+b)
- False Alarm Ratio: b/(a+b)
- Bias: (a+b)/(a+c)

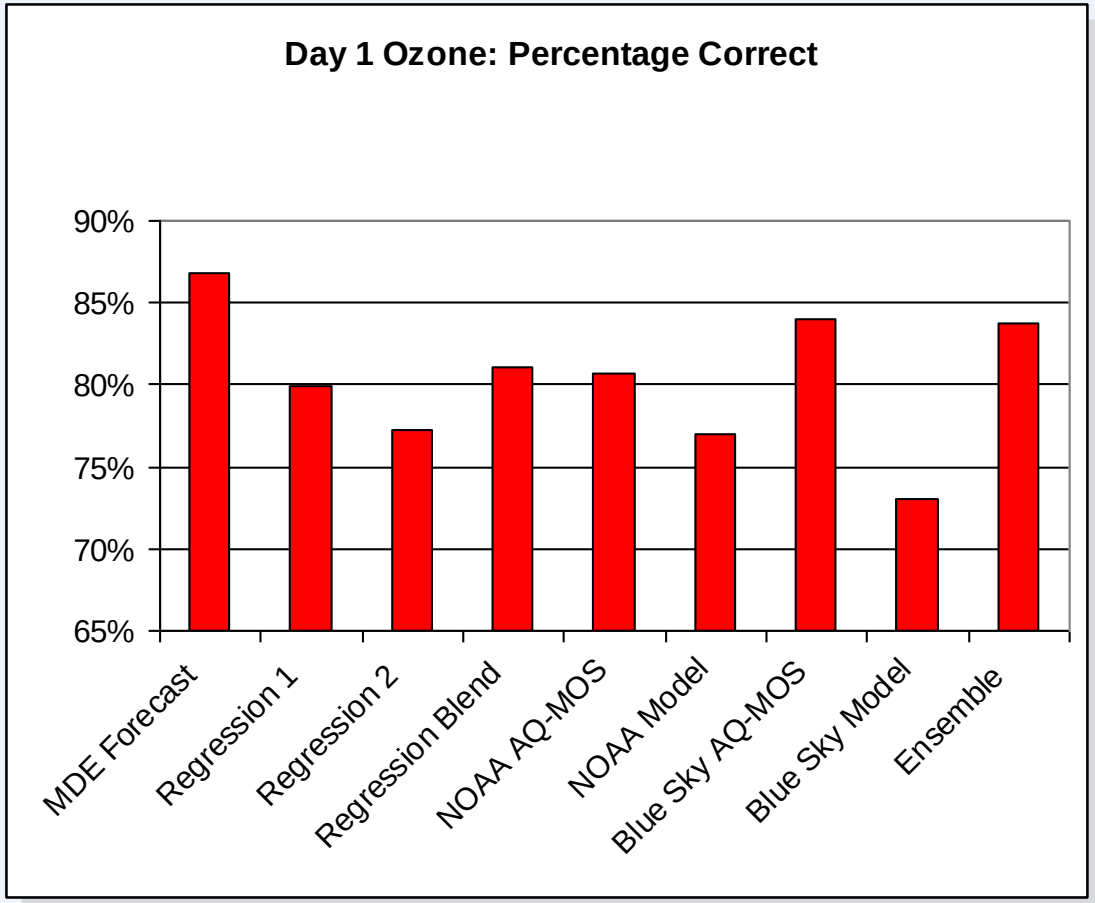
CONTINGENCY TABLE

a Forecast ≥ Y Observed ≥ Y	b Forecast ≥ Y Observed < Y
c Forecast < Y Observed ≥ Y	d Forecast < Y Observed < Y

Not comparable to last year's results because new threshold is USG (Orange) and above.

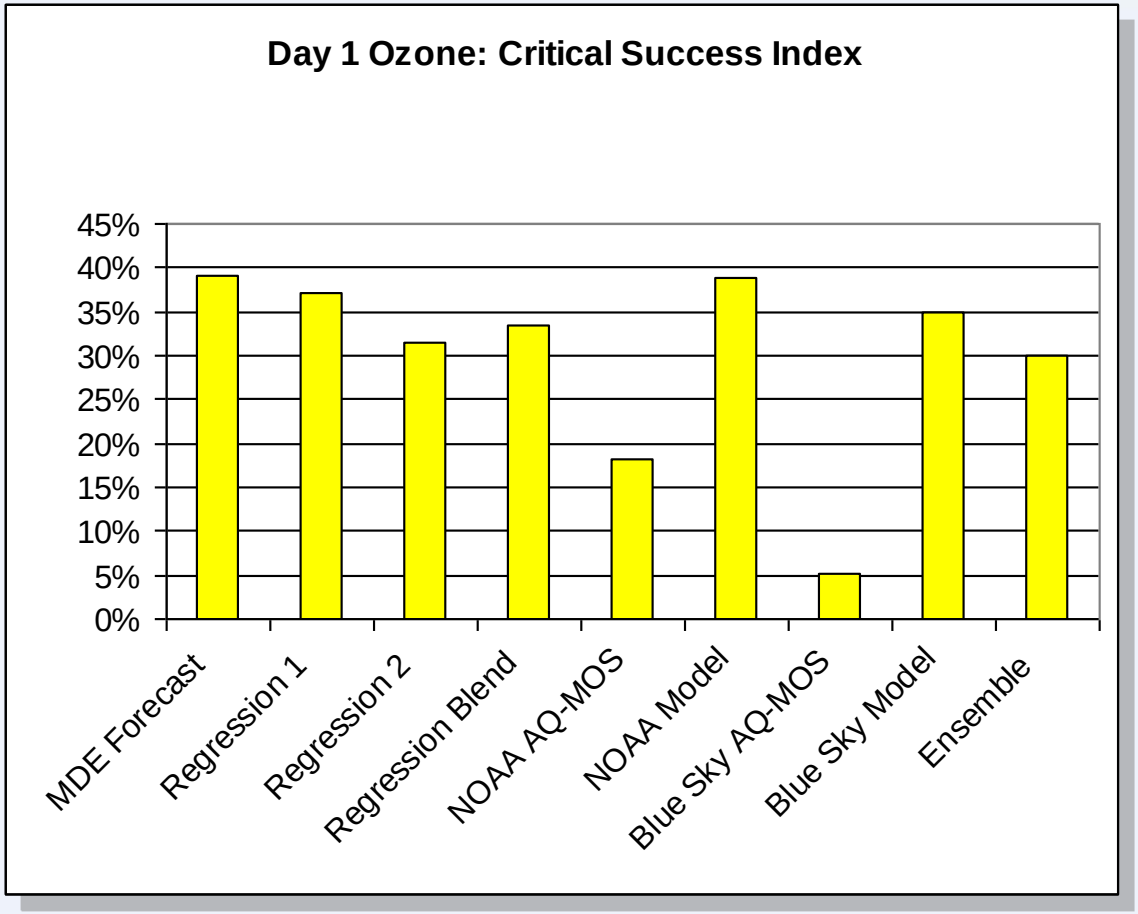
Day 1 Ozone: Percentage Correct

- MDE Forecasters scored the highest with 87% correct.
- AQ-MOS was better than raw model results.
- Ensemble did much better, given low AQ-MOS to offset high model results.



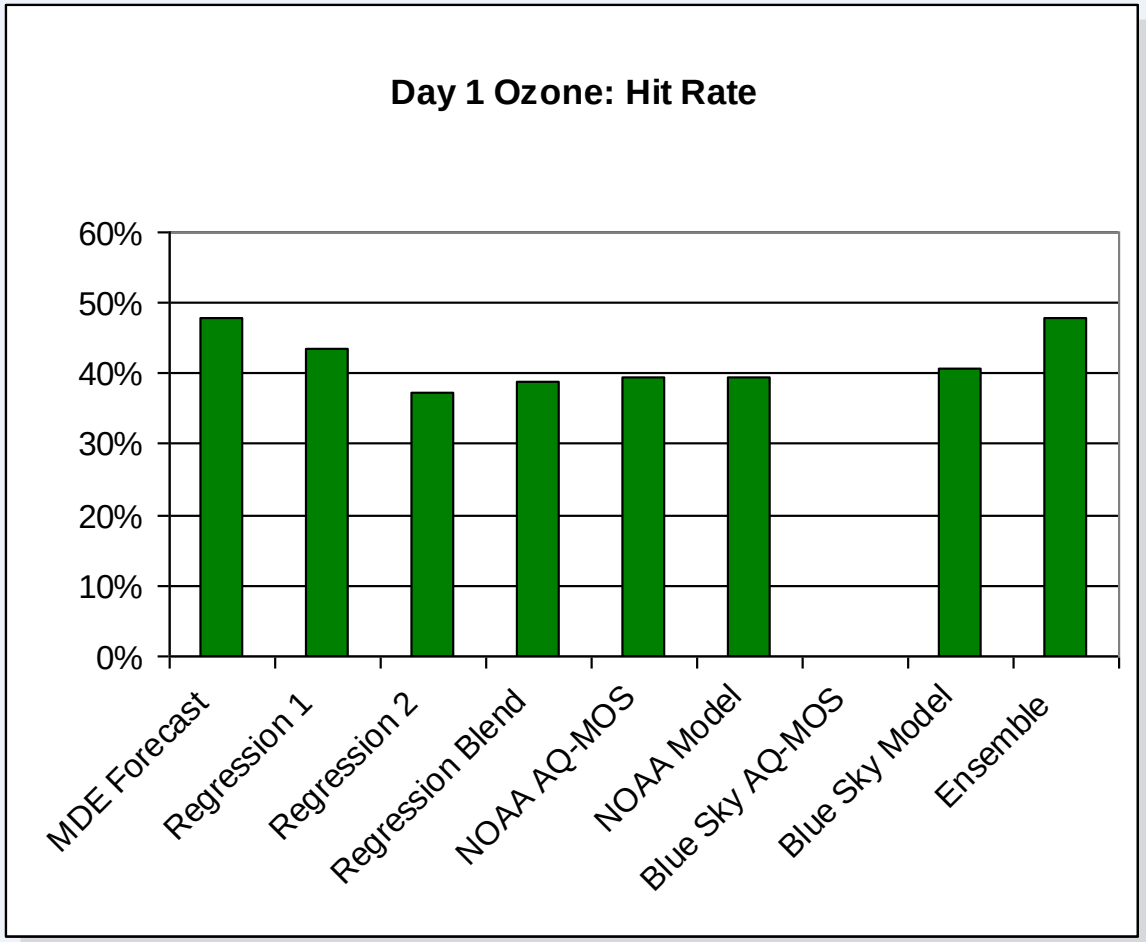
Day 1 Ozone: Critical Success Index

- MDE Forecasters and NOAA model scored the highest with 39%.
- The model results outperformed AQ-MOS.
- The Regression models also did well.



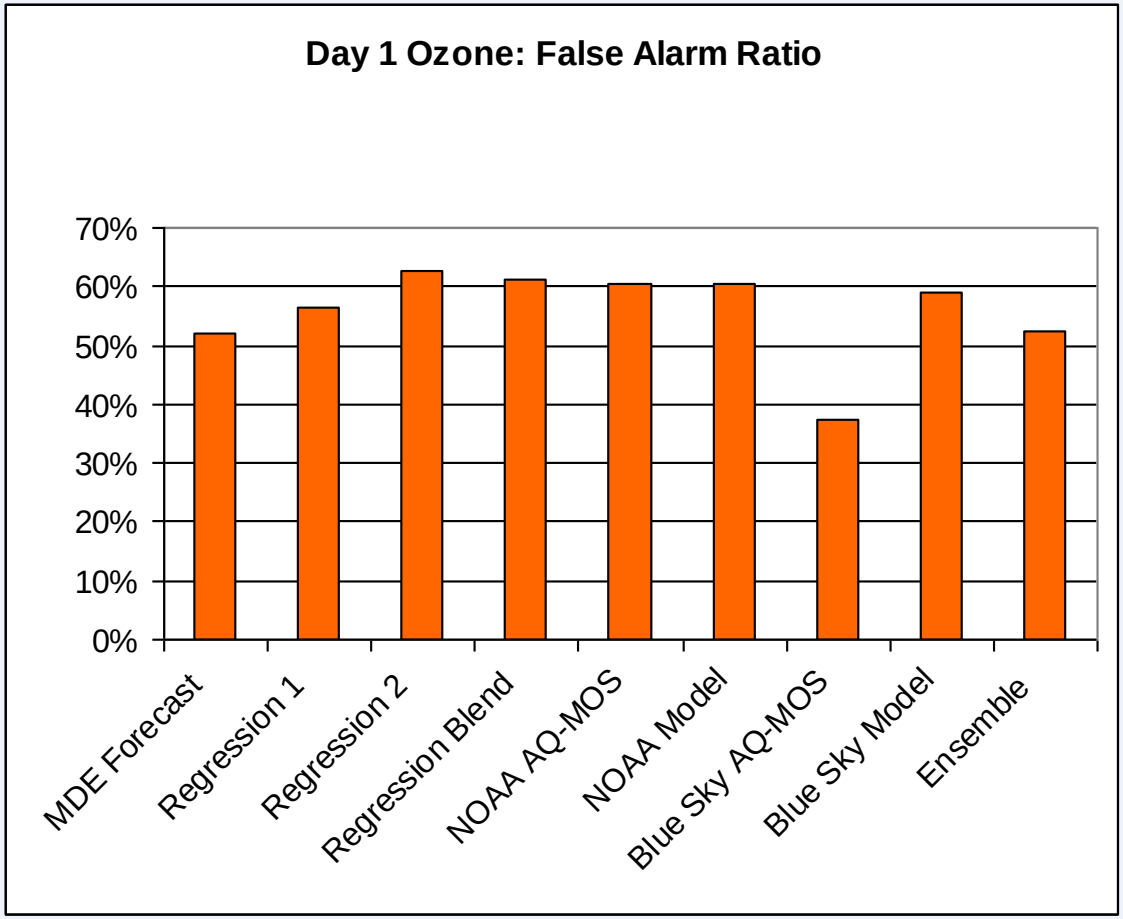
Day 1 Ozone: Hit Rate

- MDE Forecasters and the Ensemble scored the highest with 48%.
- There was little difference between NOAA and AQ-MOS.
- Blue Sky AQ-MOS had a 0% Hit Rate



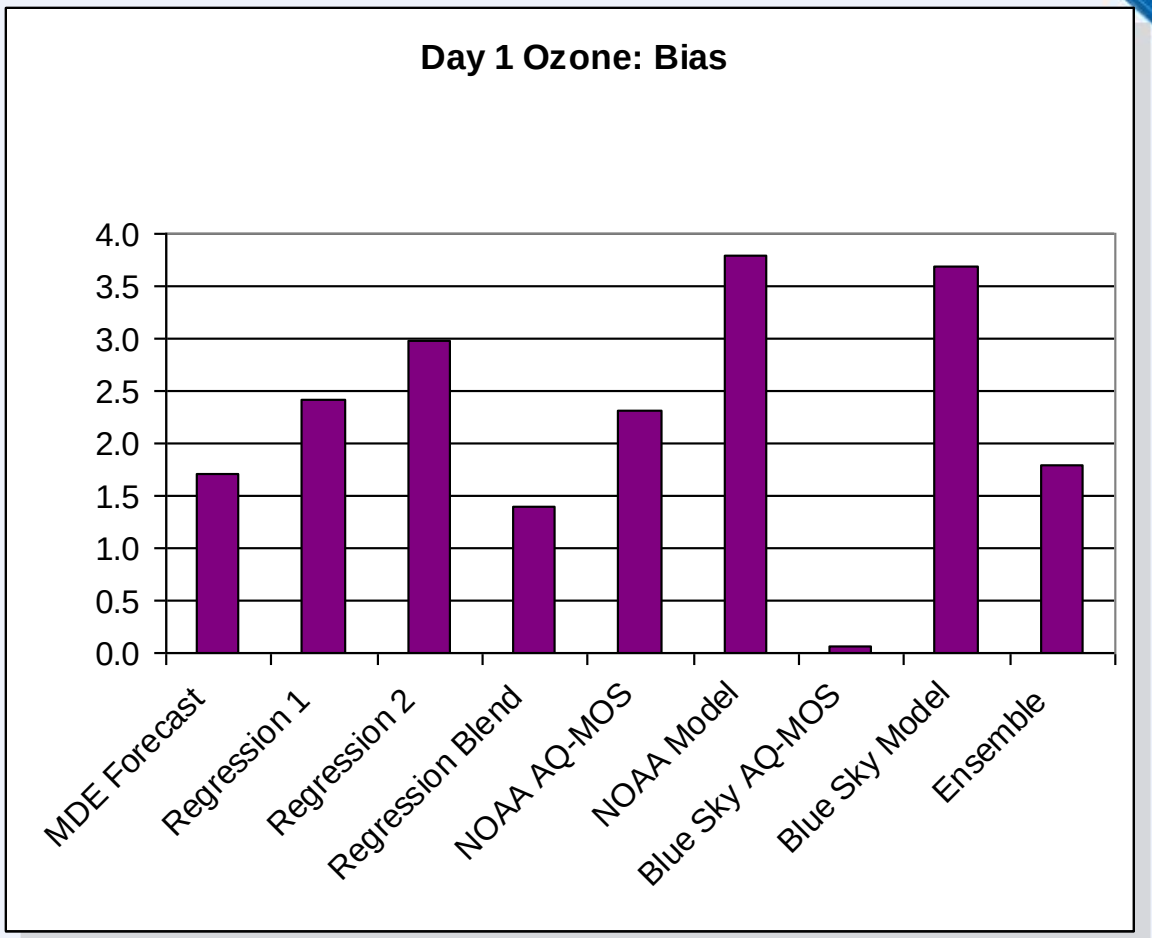
Day 1 Ozone: False Alarm Ratio

- Blue Sky AQ-MOS had the lowest false alarm rate with 38%.
- The MDE Forecasters and the Ensemble were next with 58%.
- Regression 2 had the worst ratio of 63%. The rest were also high.



Day 1 Ozone: Bias

- Blue Sky AQ-MOS was bias free.
- All others over-predicted Code Orange and above AQI.
- The highest bias was for the NOAA and Blue Sky Models.



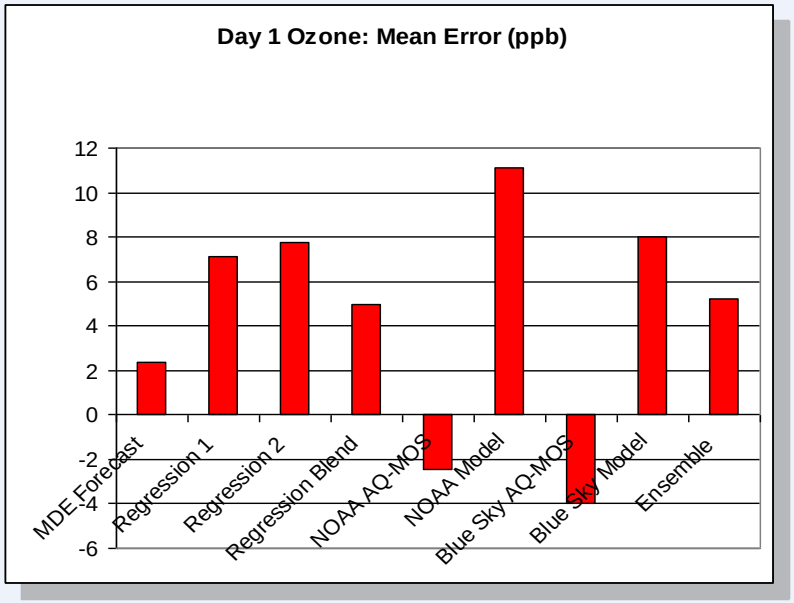
Concentration Accuracy Tests

- **Mean Error**: average of each forecast minus the paired observation.
- **Mean Square Error**: sum of squared errors divided by number of forecasts.
- **Gross Error**: summed absolute value of errors divided by the average times the total number.
- **Root Mean Square Error**: average of the square root of mean square error.

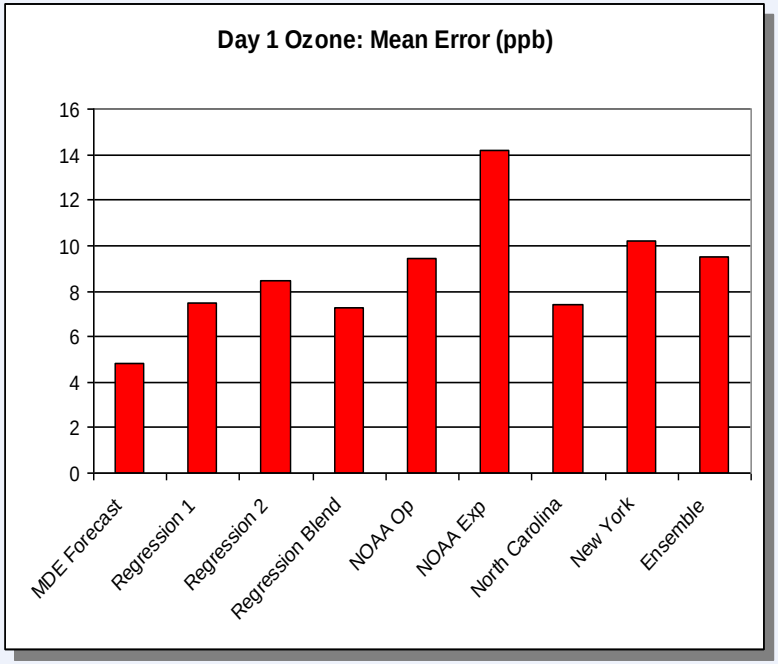
Comparable to last year's results.

Day 1 Ozone: Mean Errors

2010



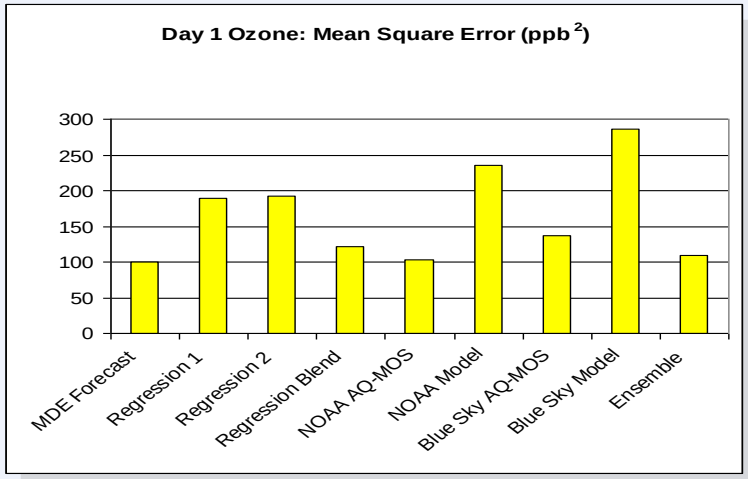
2009



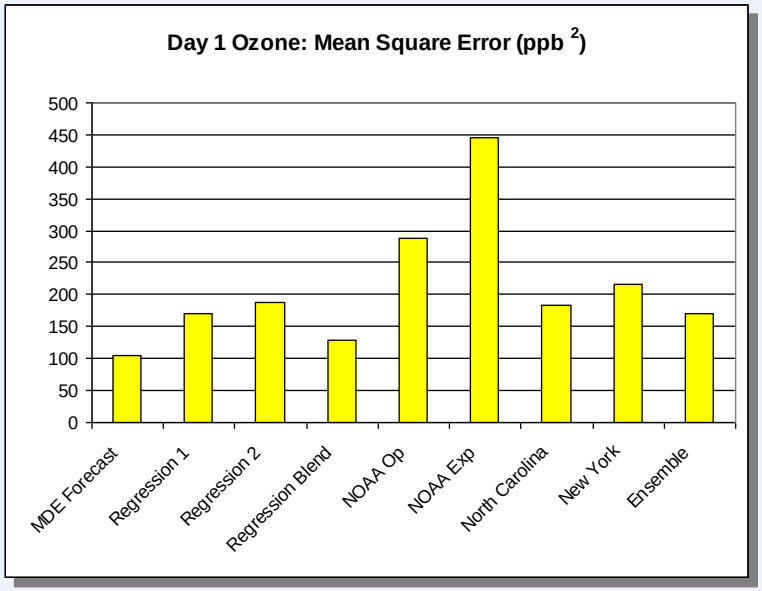
- Across the board improvements over last year!
- Both AQ-MOS have negative mean errors
- Best are MDE Forecasters and NOAA AQ-MOS (±2.4 ppb)

Day 1 Ozone: Mean Square Errors

2010



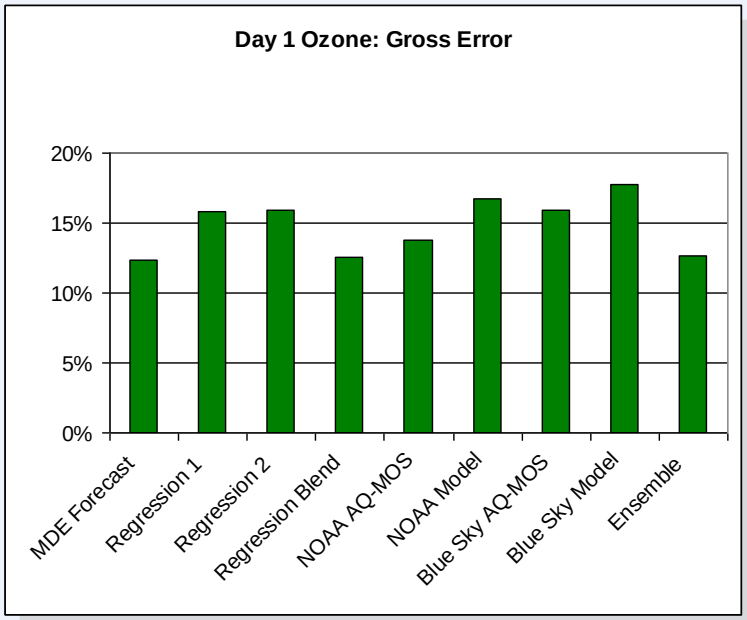
2009



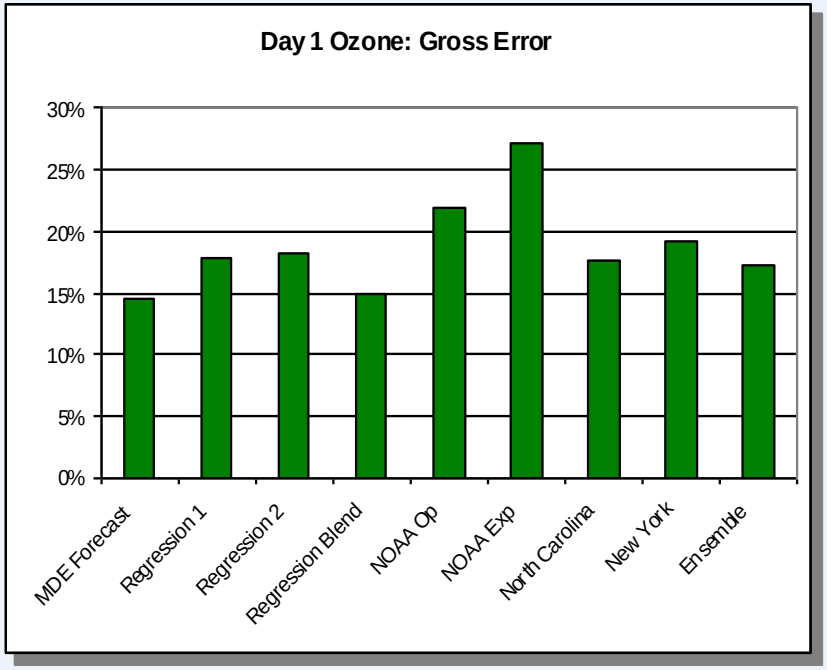
- 2010 mean square error similar to 2009 results.
- Ensemble performance improved in 2010.
- Both AQ-MOS are better than the models.

Day 1 Ozone: Gross Errors

2010



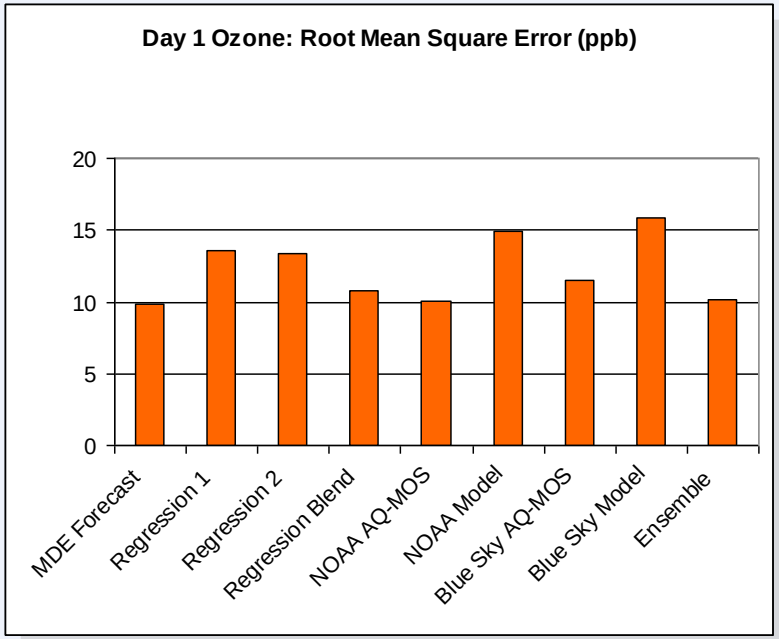
2009



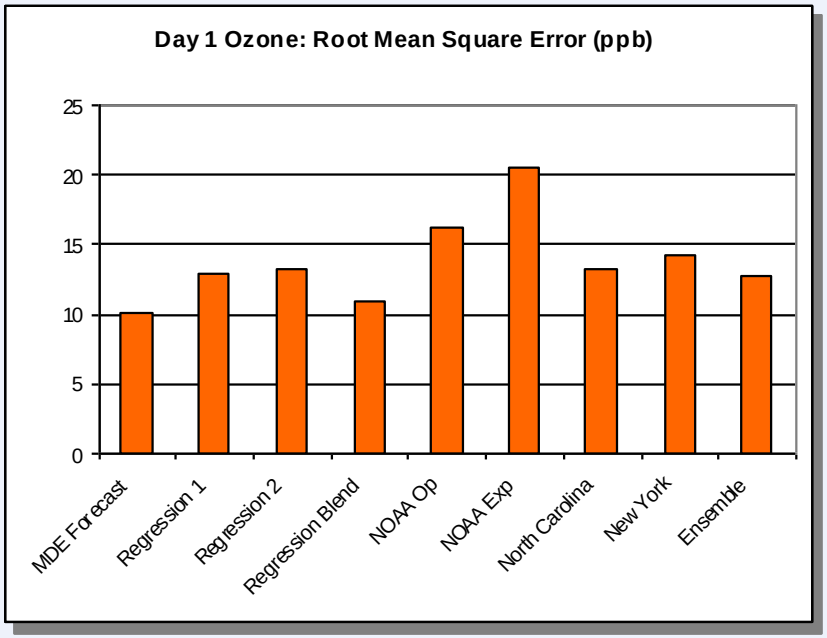
- 2010 gross errors similar to 2009 results.
- NOAA and Ensemble performance improved in 2010.
- Both AQ-MOS gross errors are similar to the models.

Day 1 Ozone: RMSE

2010



2009



- 2010 gross errors similar to 2009 results.
- NOAA and Ensemble performance improved in 2010.
- Both AQ-MOS gross errors are less than the models.

Day 1 Ozone: Ranking

1. MDE Forecasters
2. Ensemble
3. NOAA AQ-MOS
4. Blue Sky AQ-MOS
5. Regression Blend
6. Regression 1*
7. Regression 2*
8. NOAA Model
9. Blue Sky Model

* tied

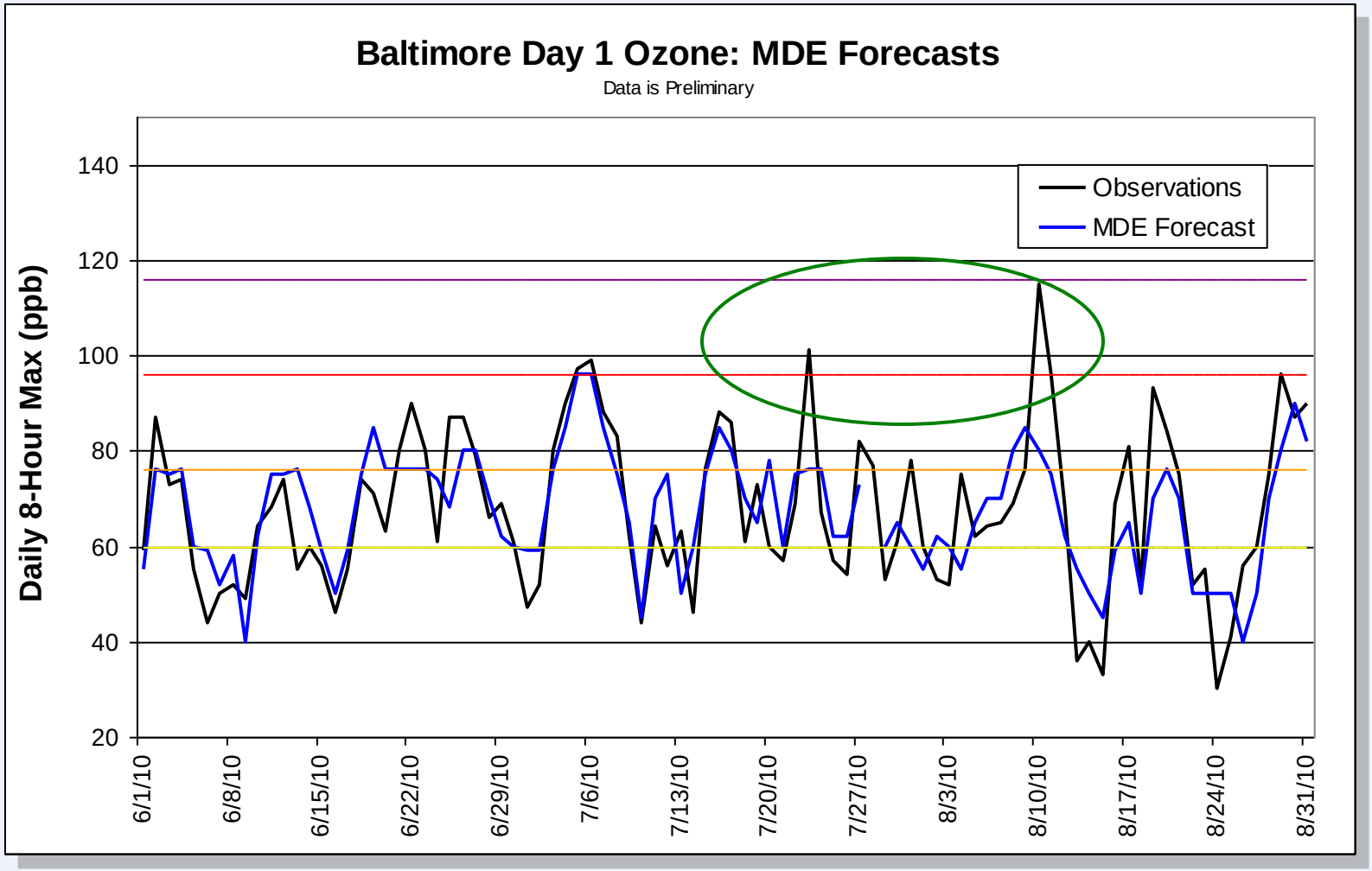
Time Series Analysis

- How does the regional performance of specific forecast tools vary from day to day?
- Were there specific days when all forecast tools performed well in a particular region?
- Were there specific days when all forecast tools performed poorly in a particular region?

NOTE: All time series are for daily 8-hour maximum ozone.

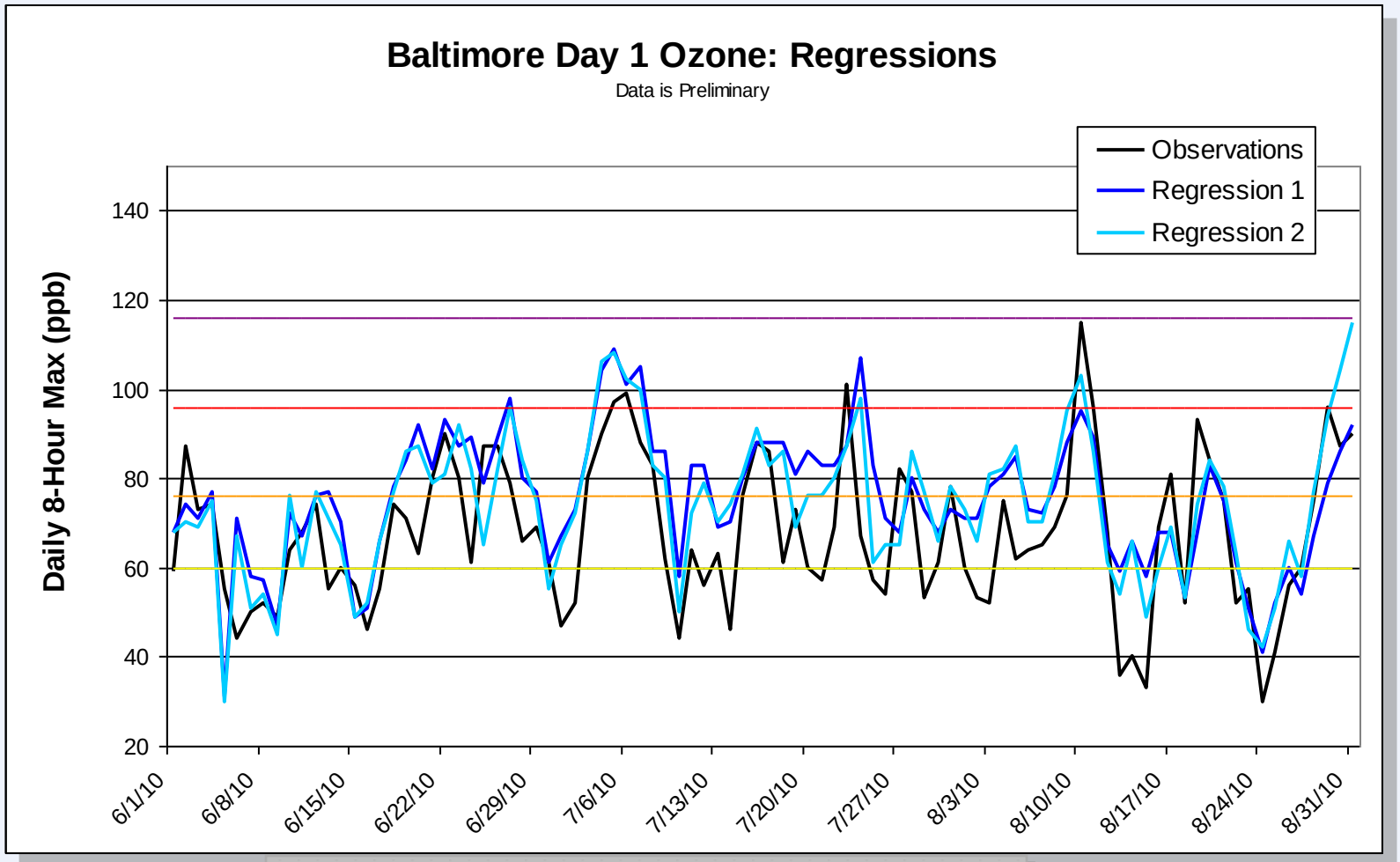
Baltimore Region Ozone Forecasts

MDE Forecast closest to observations.



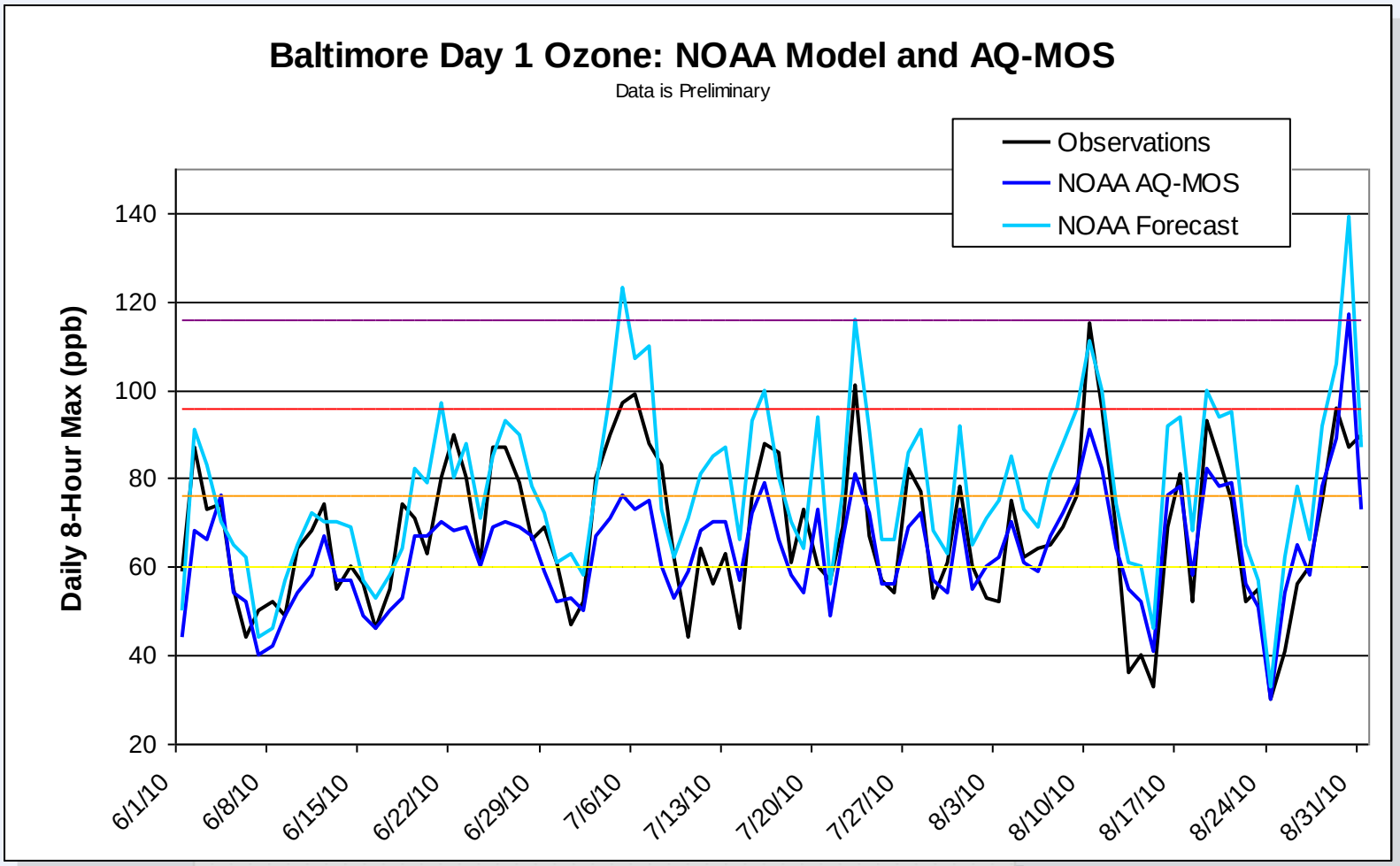
We had problems forecasting days above Code Orange.

Regressions were overpredicting.



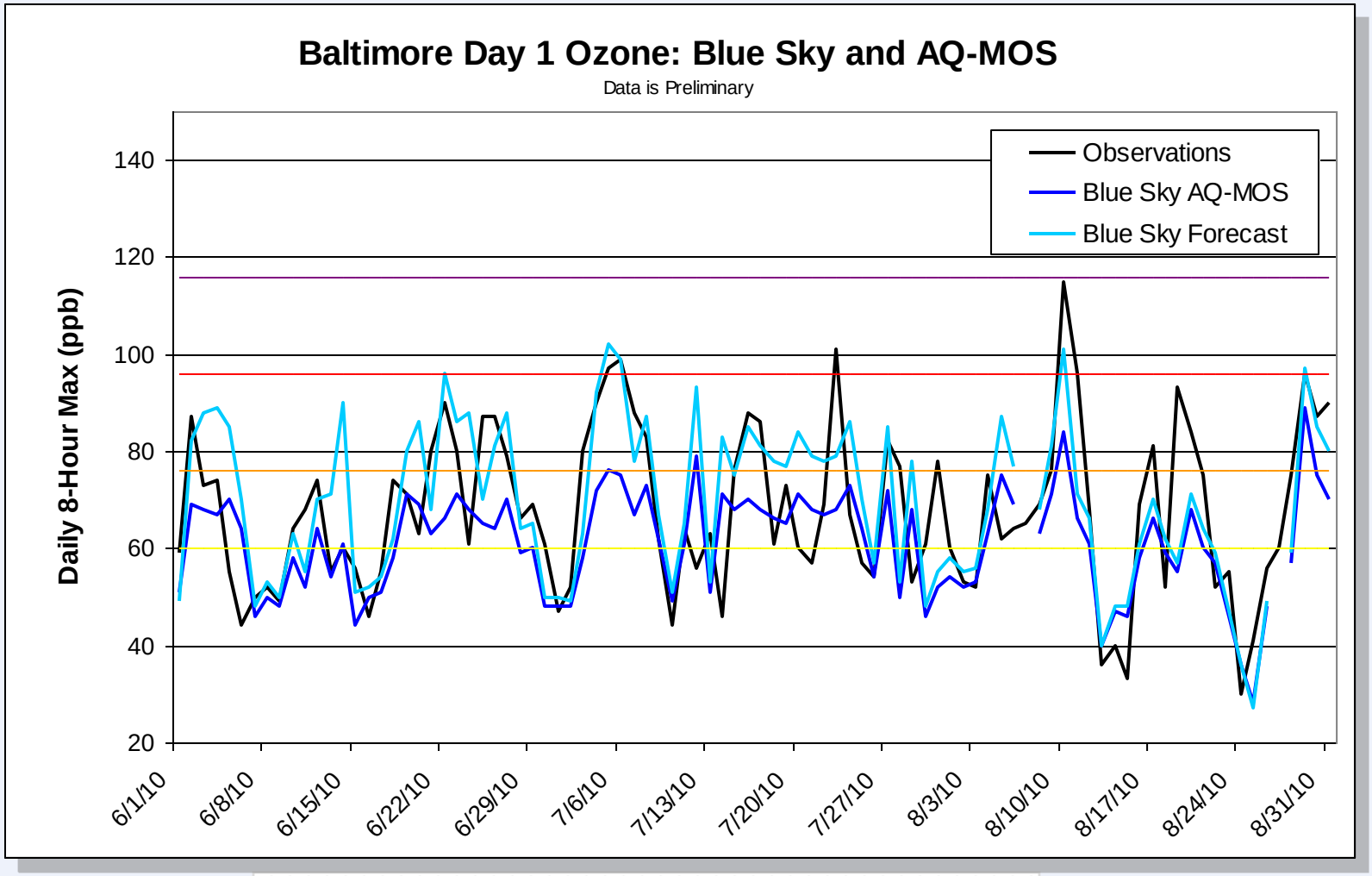
Best results were in August.

NOAA model overpredicts ozone.



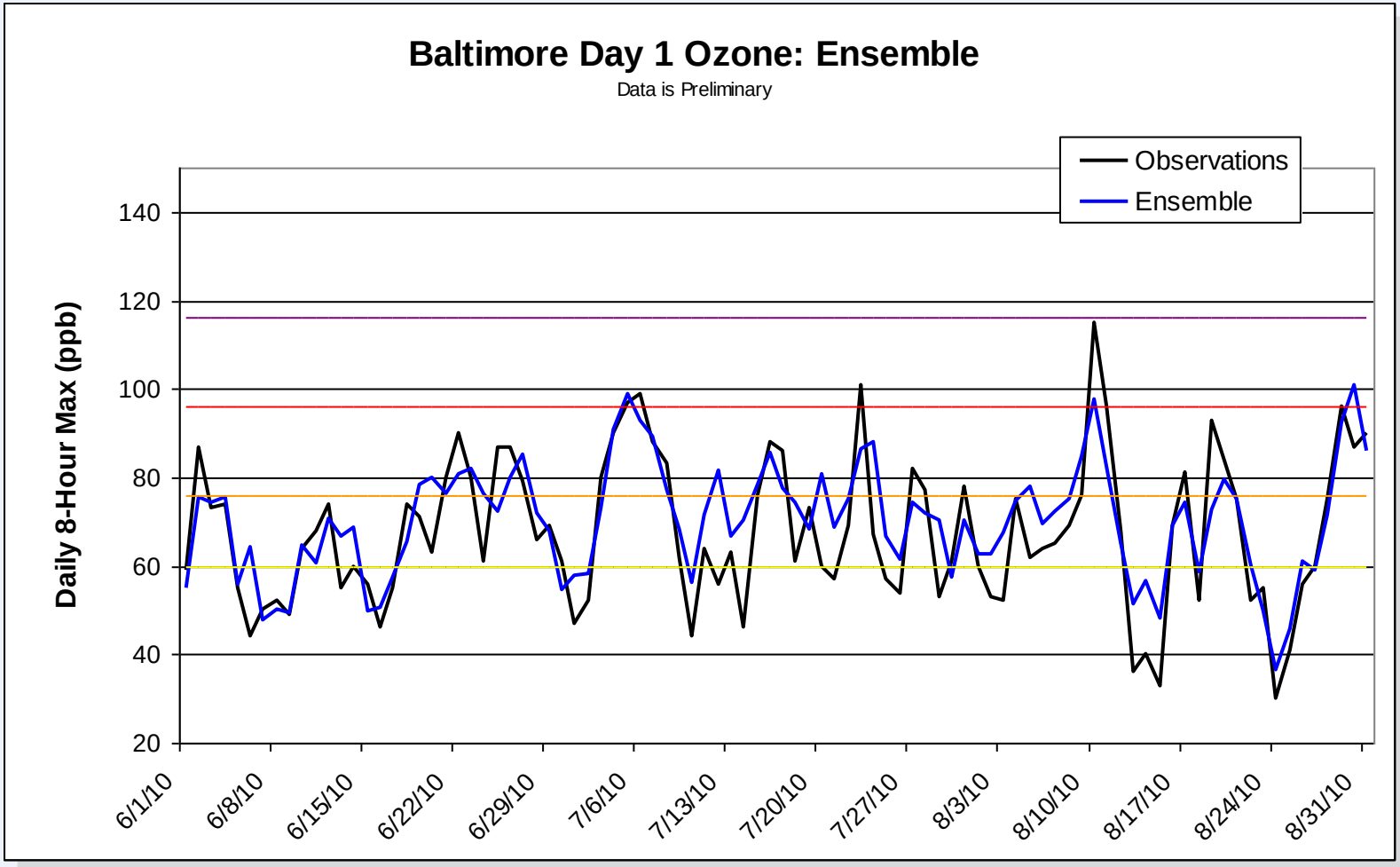
AQ-MOS underpredicts ozone.

Blue Sky overpredicted even more.



Again, AQ-MOS underpredicted.

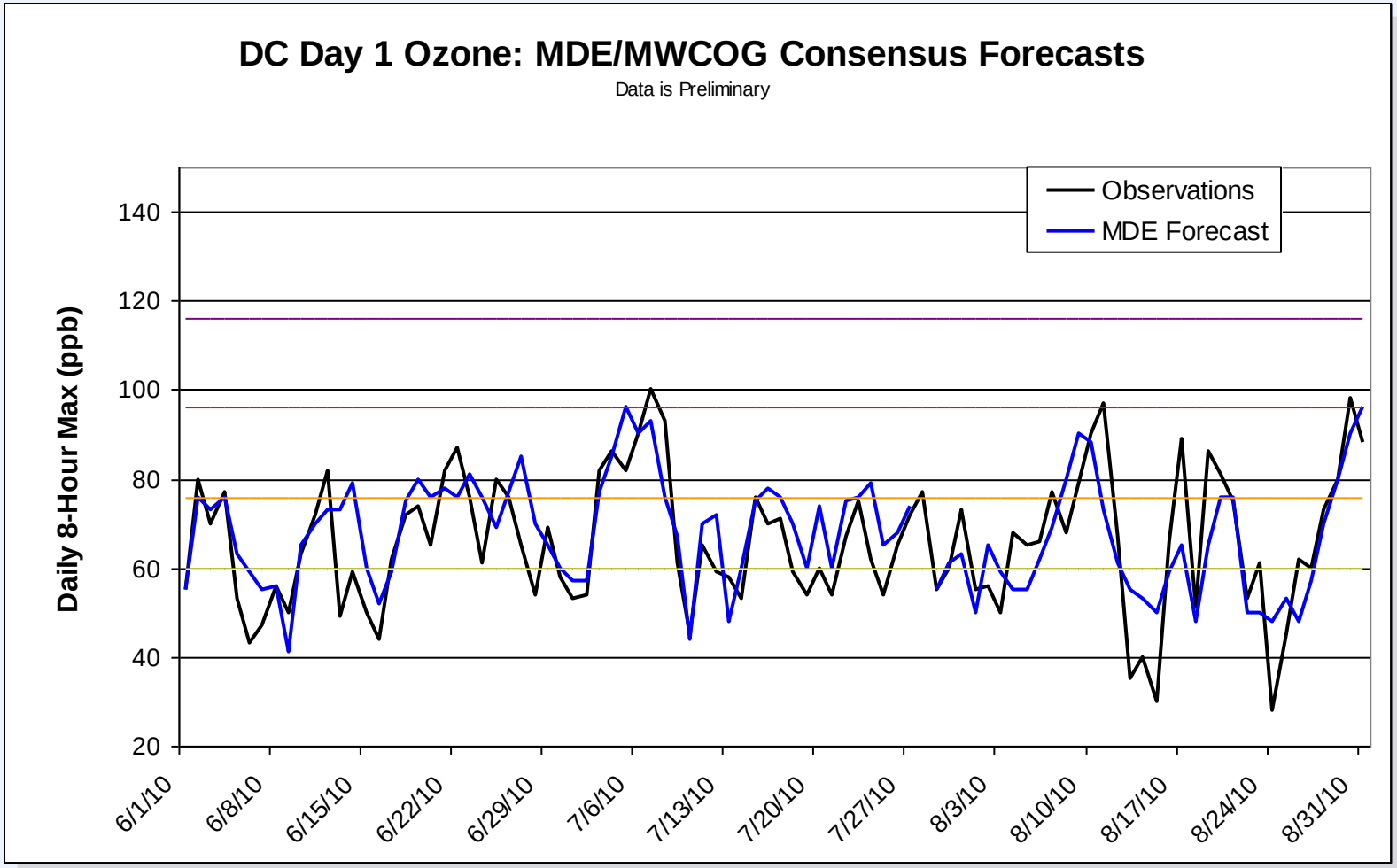
Ensemble better than models.



Big improvements over last year.

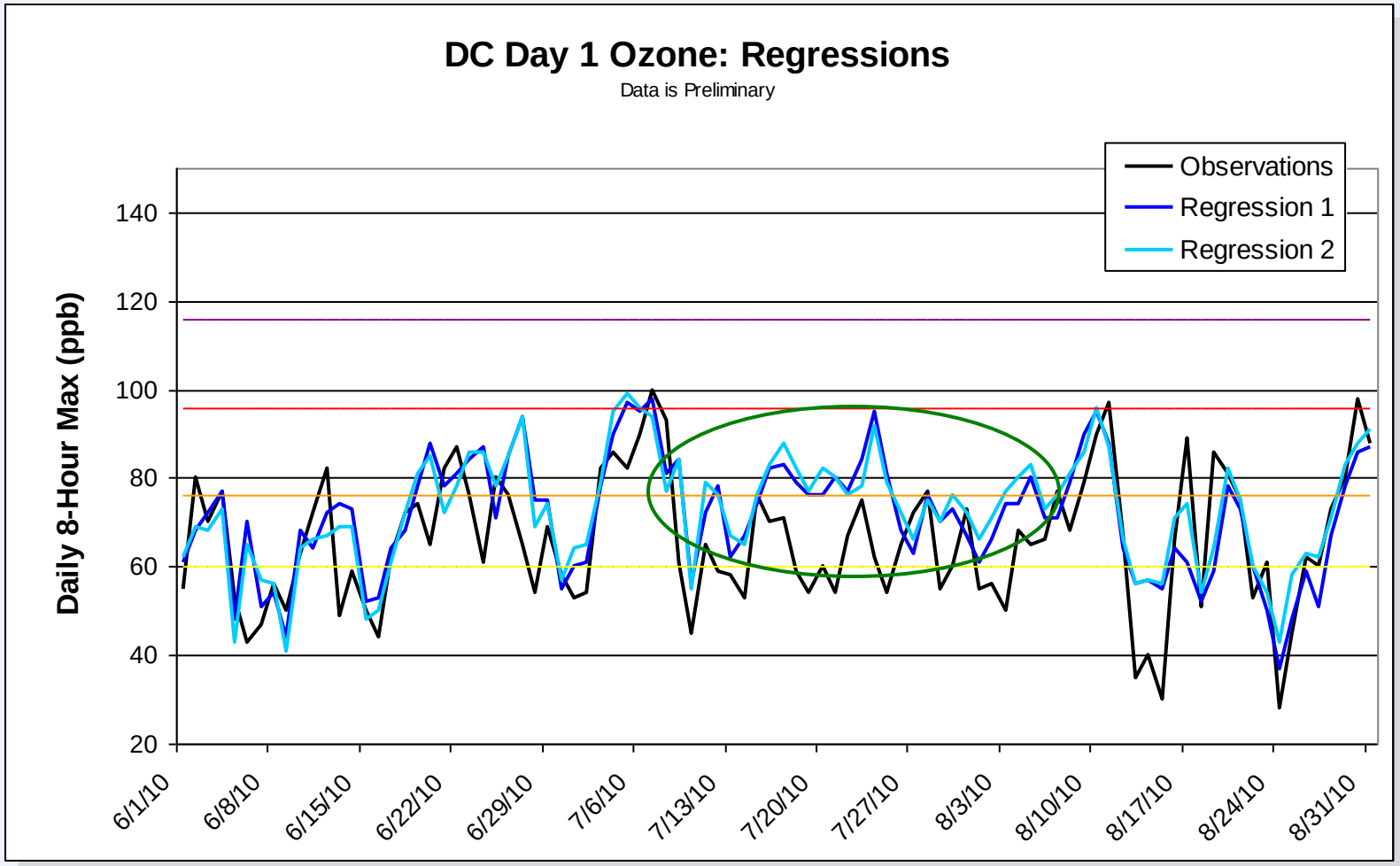
DC Metropolitan Region Ozone Forecasts

MDE Forecast closest to observations.



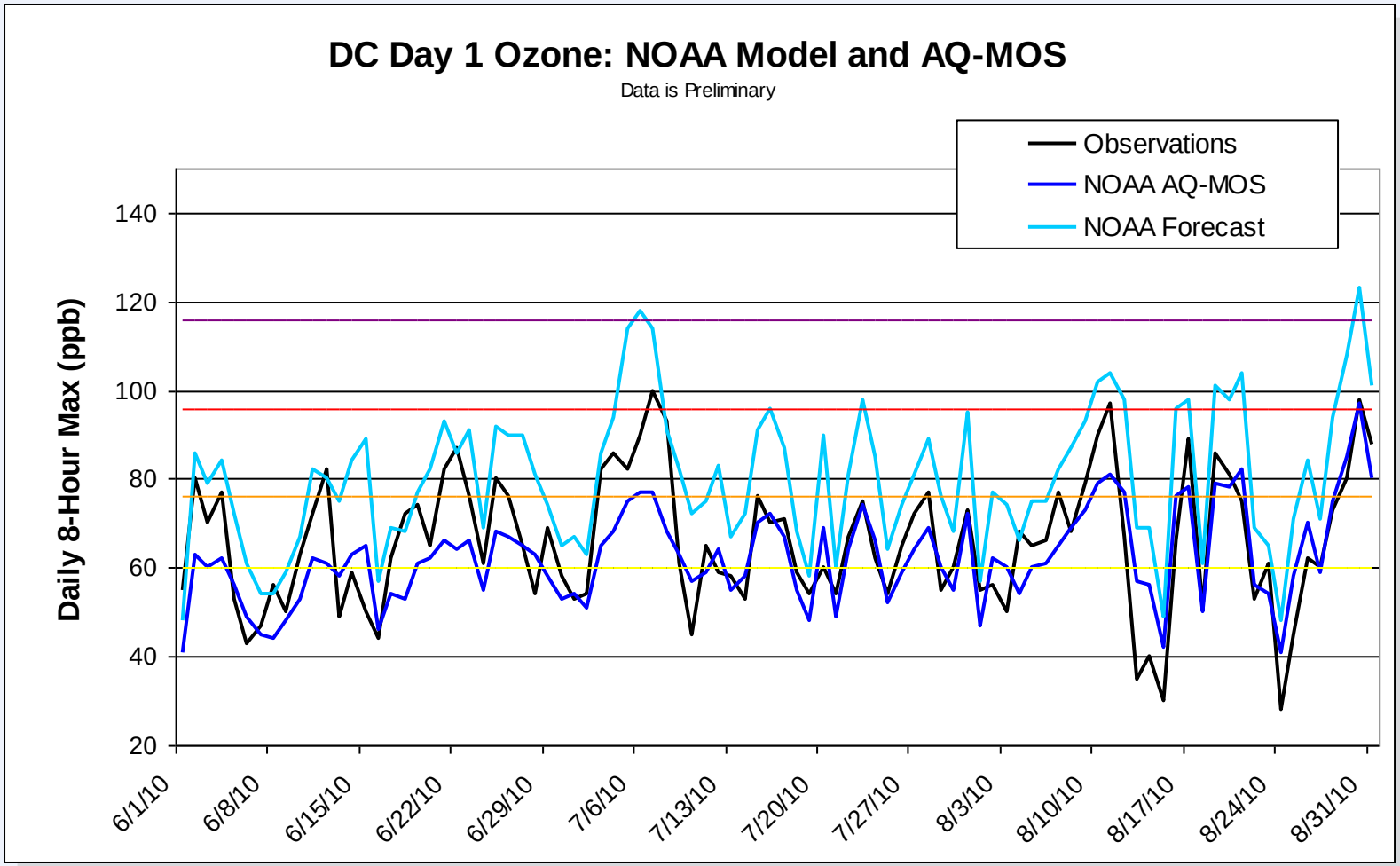
Better than forecasts for Baltimore: no high peaks.

Regressions do well overall.



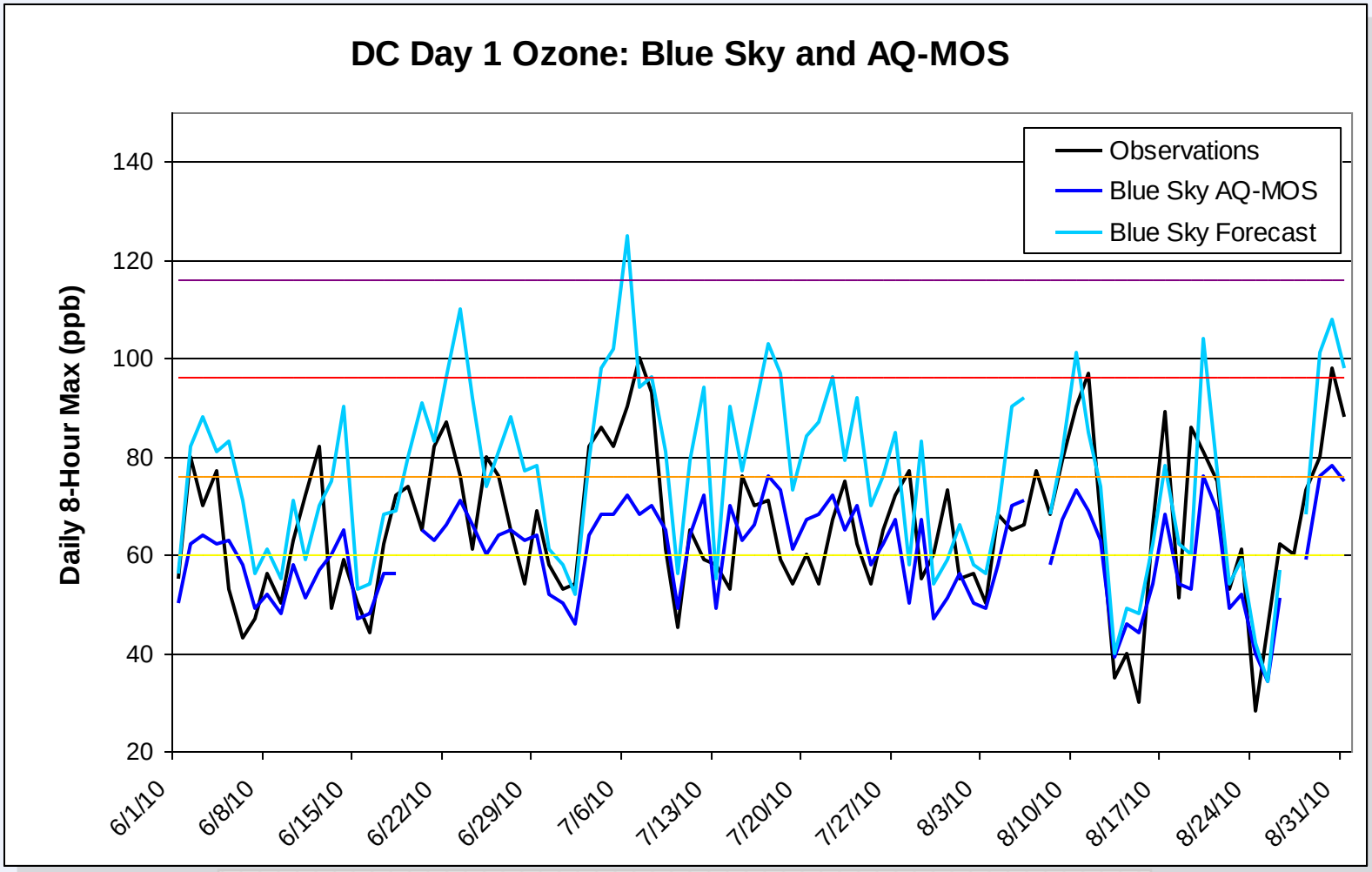
Big overpredictions in July.

NOAA model overpredicts ozone.



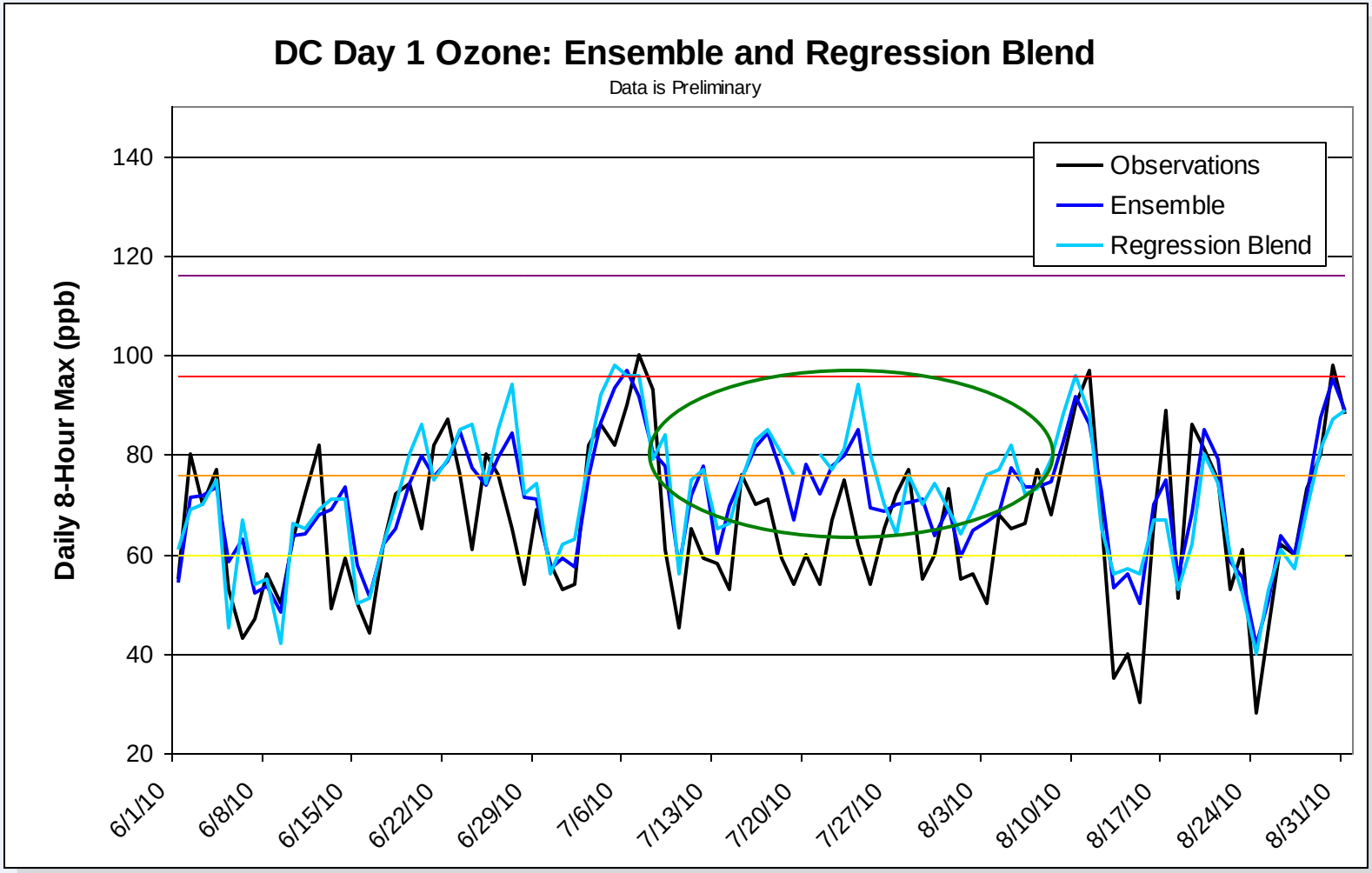
AQ-MOS underpredicts ozone.

Blue Sky overpredicts even more.



AQ-MOS underpredicts ozone.

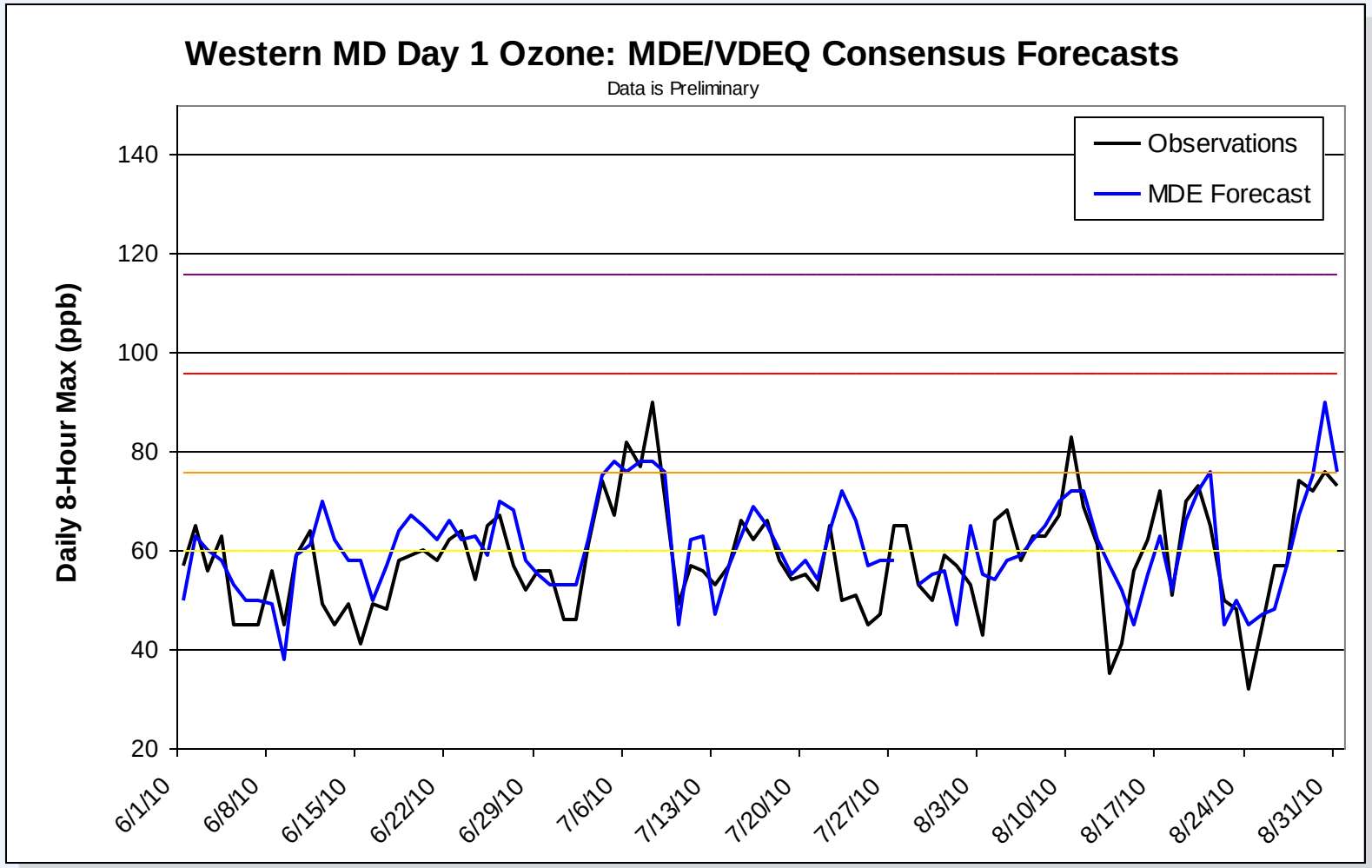
Ensemble performs well overall.



Similar but lesser problem in July.

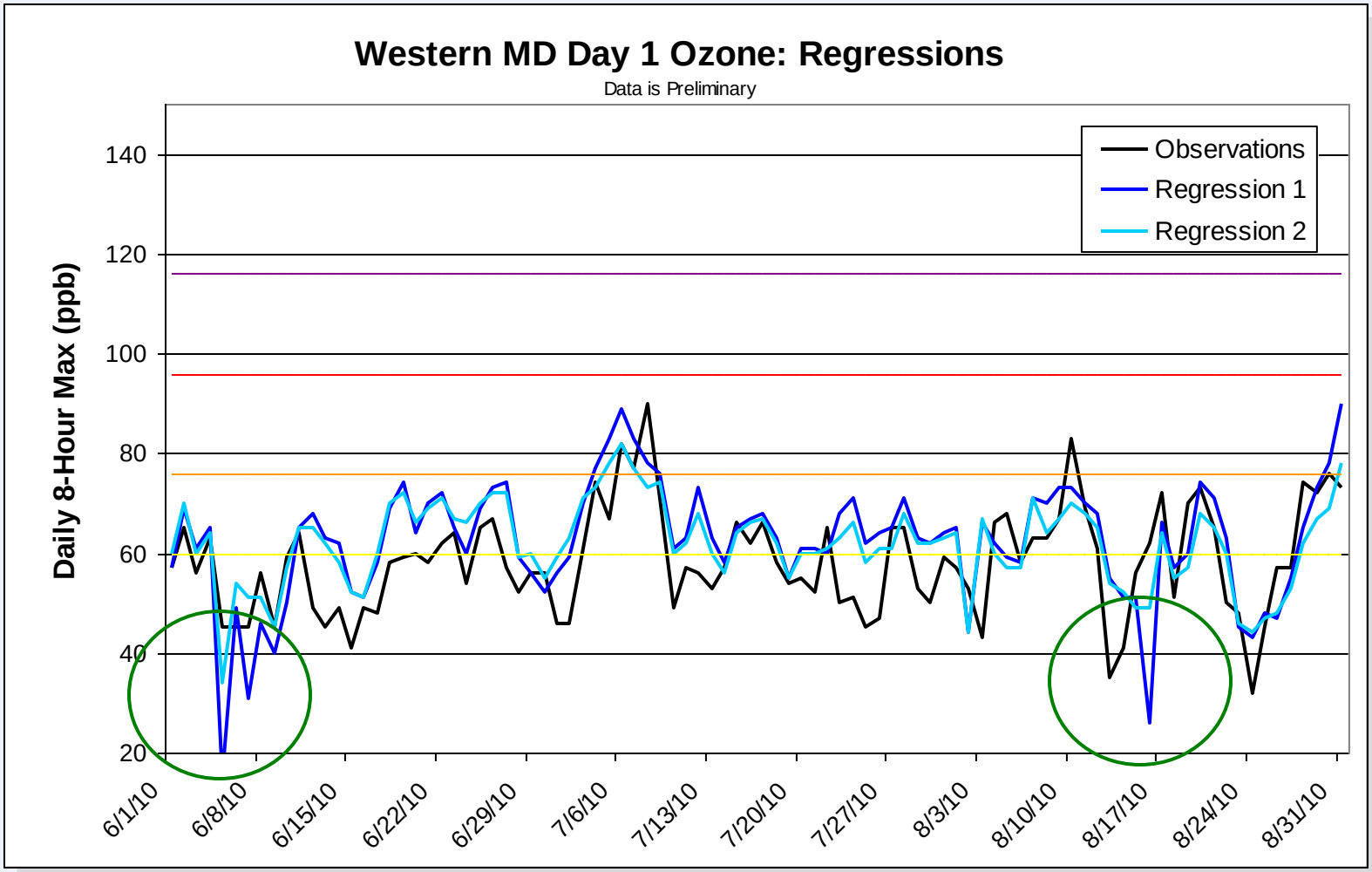
Western Maryland Region Ozone Forecasts

MDE Forecast closest to observations.



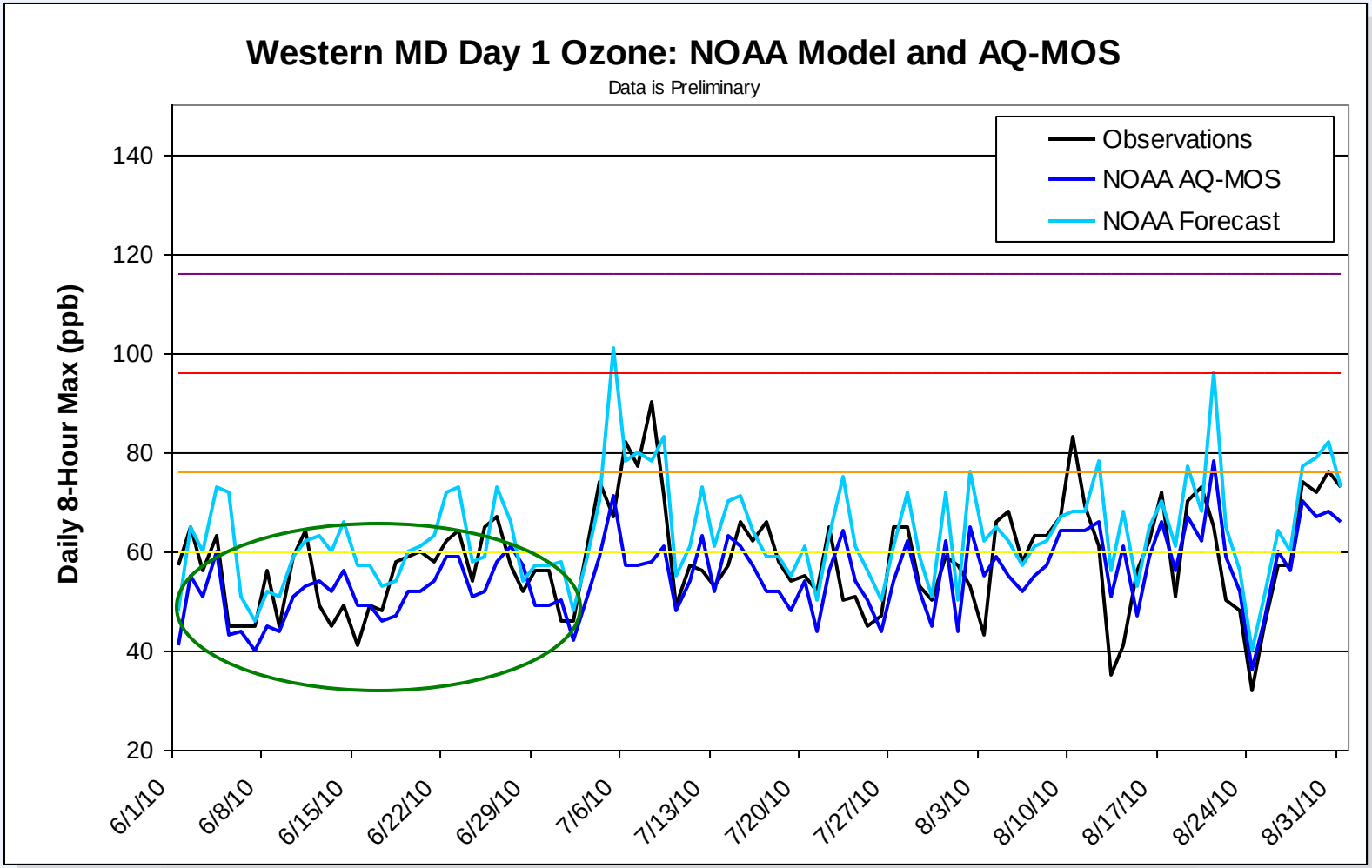
Note that only Maryland sites are included in observations.

Regressions did well in WM.



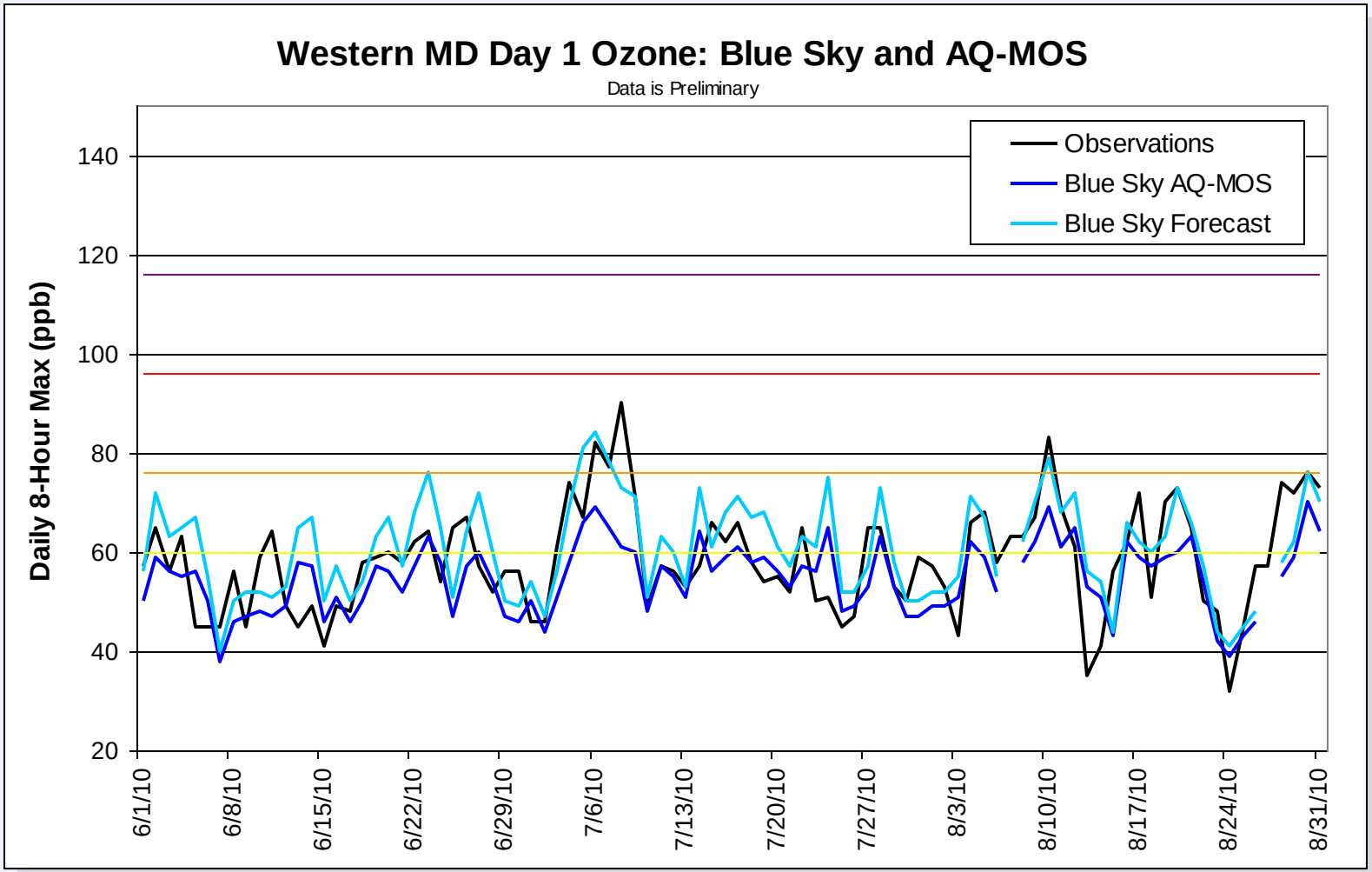
Couple of strong underpredictions.

NOAA models very good in WM.



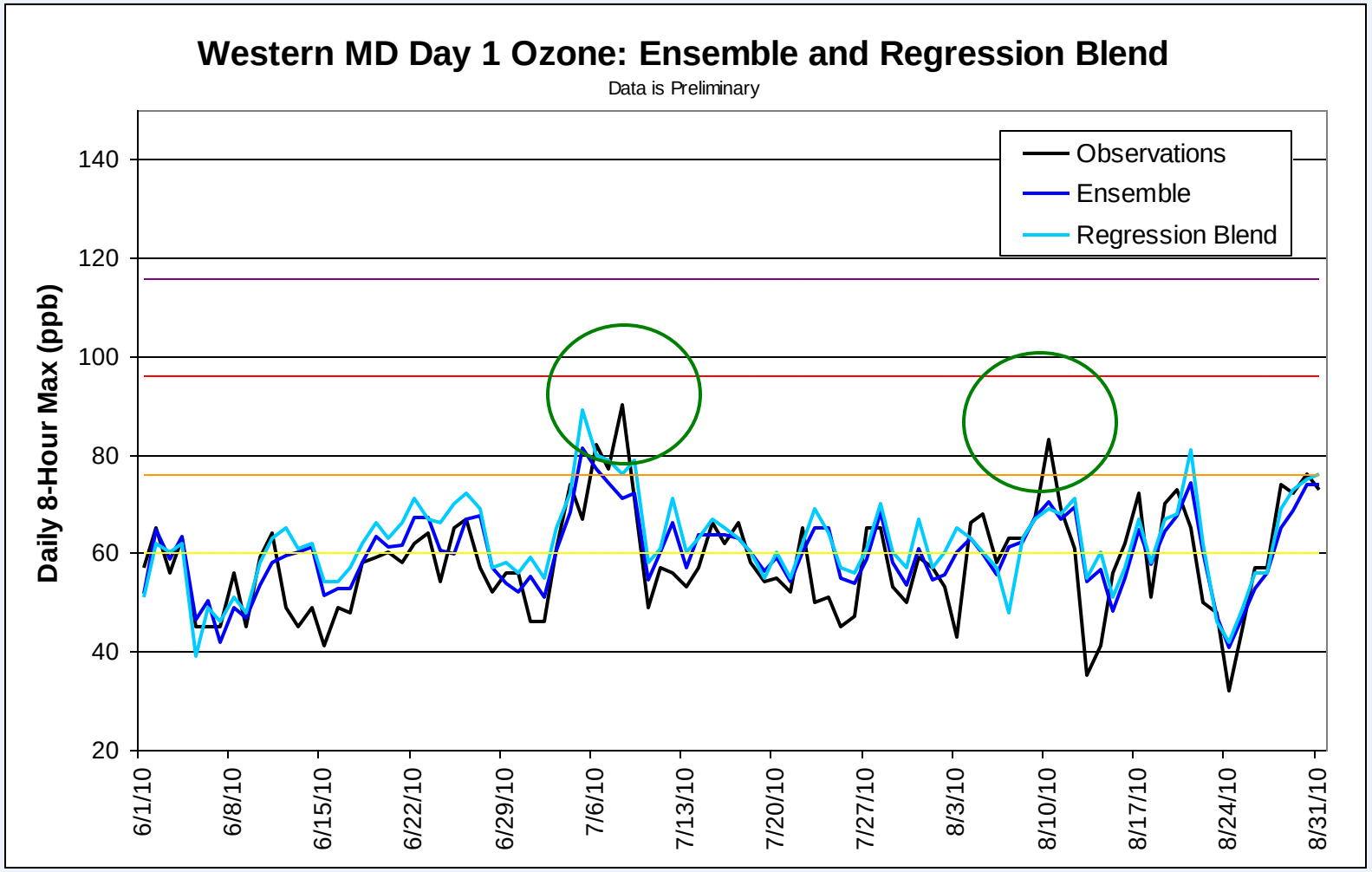
AQ-MOS undepredictions in June.

Blue Sky overpredicts ozone.



AQ-MOS does very well in WM.

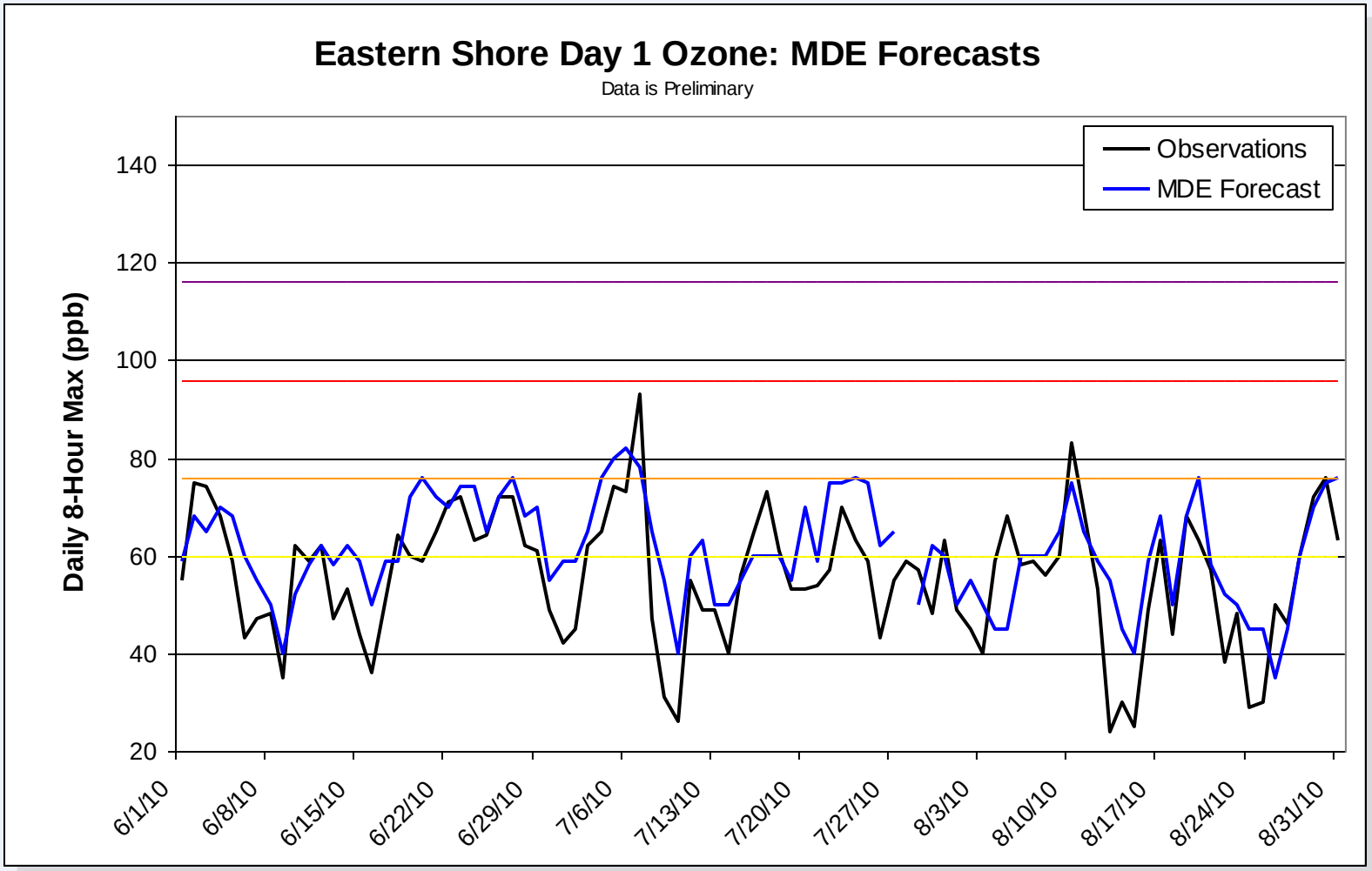
Ensemble tracks Regression Blend.



Two Code Orange peaks missed.

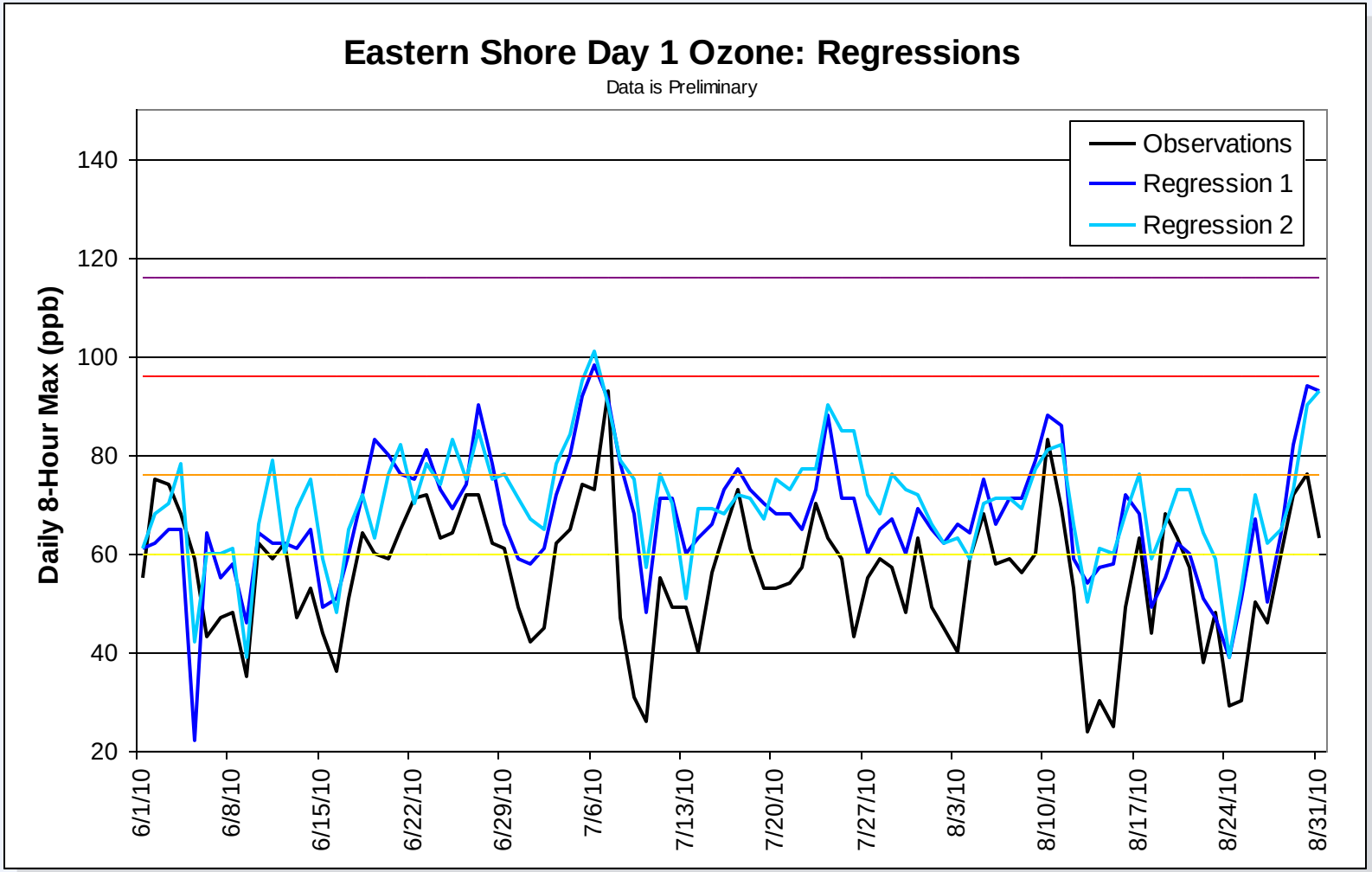
Eastern Shore Region Ozone Forecasts

MDE bests all Forecast Tools.



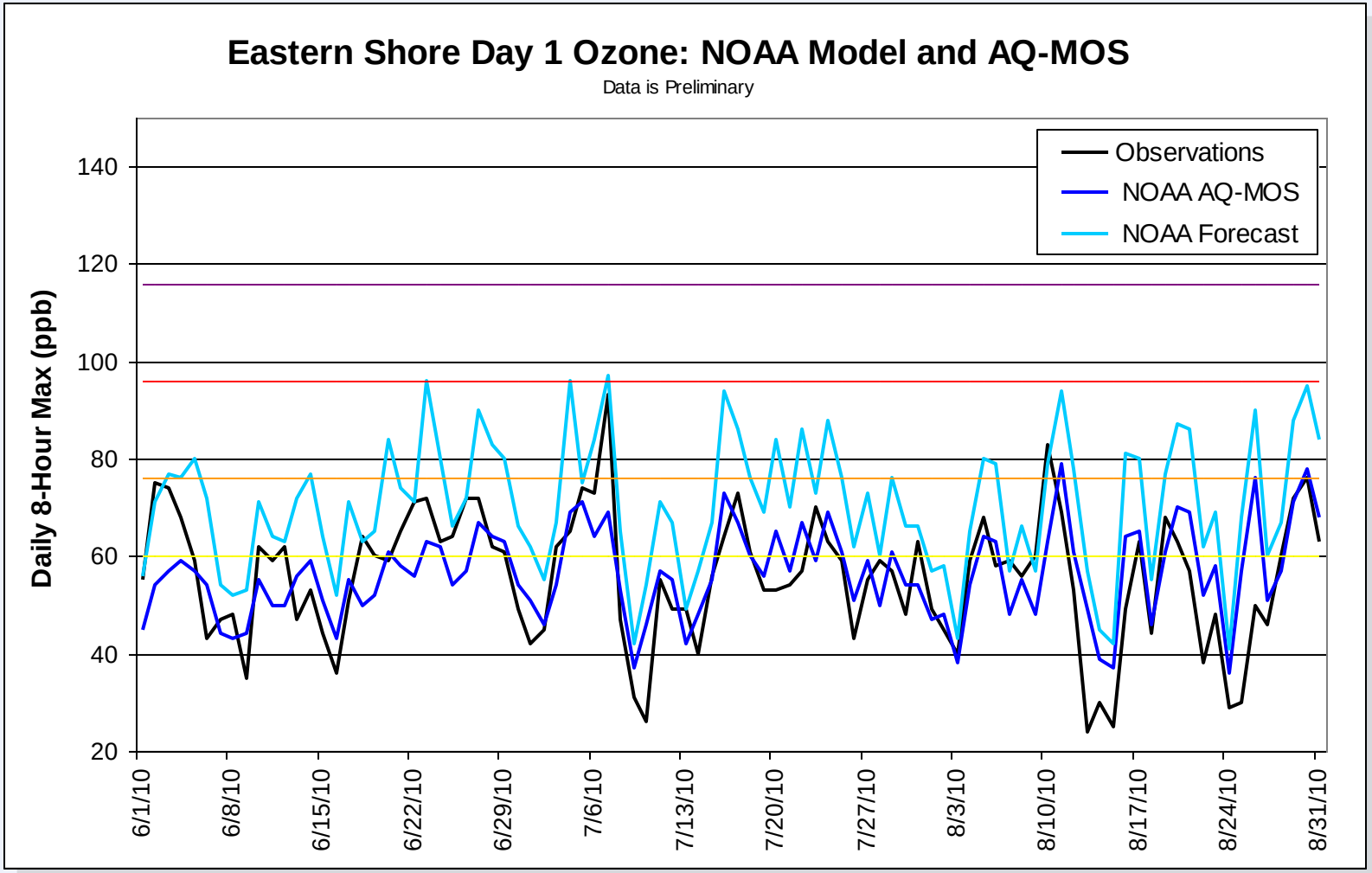
Most overpredict ozone in ES.

Regressions overpredict ozone.



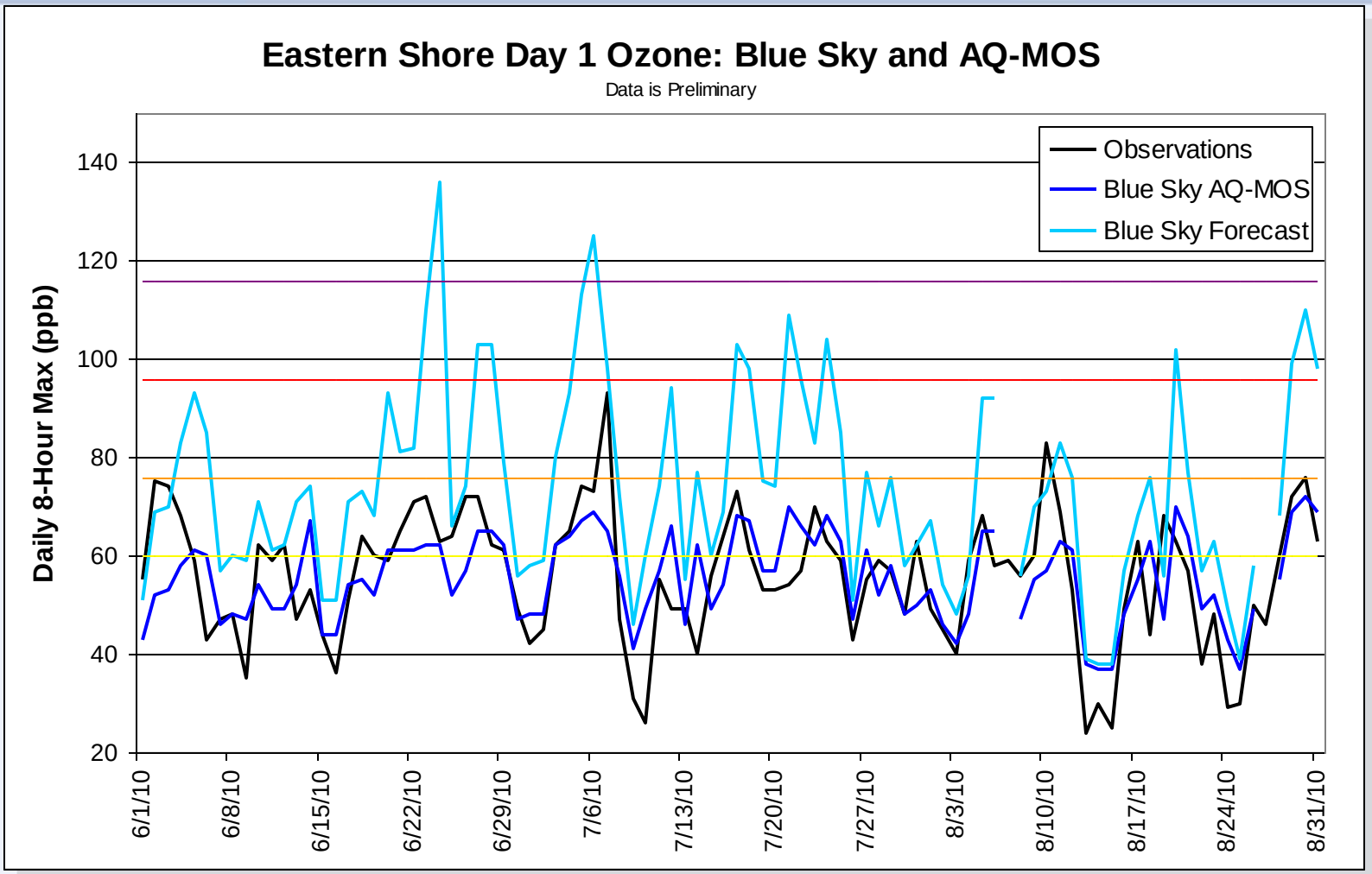
Regression 1 is slightly better than Regression 2.

NOAA model also overpredict.



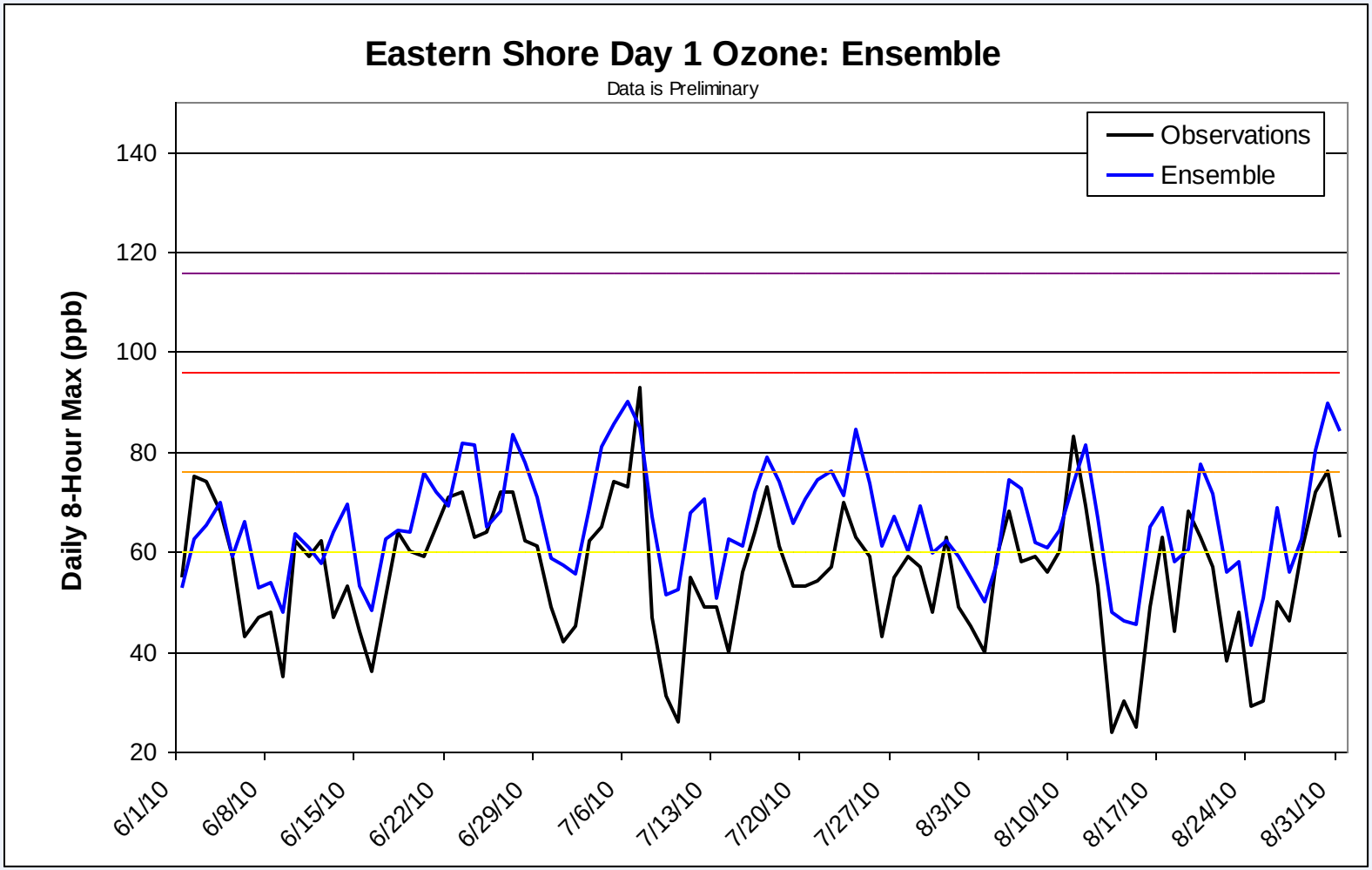
Even AQ-MOS has a number of overpredictions.

Blue Sky has trouble in ES.



AQ-MOS does well again.

Ensemble not as unreliable for ES.



Models and regressions overpredict ozone.

Forecast Recommendations

- Use NOAA/Blue Sky Model results as maximums.
- Use both AQ-MOS as minimums.
- Then forecast within this range.
- Use Ensemble for more precise guidance.
- Encourage multi-day runs for AQ forecast models (NC Model: Please come back!!)

Contact

Debra Baker
Air Monitoring Program

Air & Radiation Management Administration
Maryland Department of the Environment

Dbaker@mde.state.md.us