We thank the reviewers for their constructive comments. We have addressed all of them and made changes to the manuscript, indicated by italics.

Reviewer #1:

General Comments

This paper presents four different empirical models for estimating peak daily maximum 8-hour (averaged) ozone from meteorological factors and the level of nitrogen oxide emissions (NOx = NO2 + NO). Four different statistical models: the Generalized Additive Model (GAM), the Multivariate Adaptive Regression Splines, the Random Forest, and the Support Vector Regression were developed and applied to estimate ozone concentrations in the South Coast 25 Air Basin (SoCAB) of California, including Los Angeles and the surrounding region.

The use of empirical models for estimating extreme ozone concentrations is particularly relevant because these results may be of interest for health and regulatory purposes. The models may be improved further as the available datasets become larger due to more observations being made over time. Empirical / statistical models are usually more accurate than first-principles numerical forecast models (as long as there are no large changes in the conditions used to derive the empirical models).

There is a long history of the development of empirical models that extends back several decades. However, there are new concepts in machine learning that the authors have used to inform their research. I commend the authors for their appropriate citation of the literature, but they might consider a paragraph to mention the long history of empirical / statistical models.

Specific Comments

The authors provide an excellent discussion of their four models: the Generalized Additive Model (GAM), the Multivariate Adaptive Regression Splines, the Random Forest, and the Support Vector Regression. This clearly written presentation is an outstanding introduction to modern empirical modeling. I can easily imagine using this paper in a graduate atmospheric science course.

The correlations between the model predictions and observations are high and the biases are low. There are only small differences in the performance, in terms of accuracy and required computational resources, between the four approaches for the dataset examined. It would be interesting to see a similar comparison for a much larger dataset in a future paper. Overall, I find that this paper by Gao et al. to be a valuable contribution to the literature.

Technical Corrections

Please consider a paragraph to mention the long history of empirical / statistical models if space allows.

We added the following sentences from line 34 to line 49 about the long history of statistical methods:

"However, the rise of machine learning methods, along with an increasingly long observational record, suggests that observation-based, statistical models can be used to understand those relationships. Since the 17th century statistics have been used to record information about wealth and population in Europe (Porter, 1981). For example, William Petty, a British scientist and economist, estimated the census data of Ireland through statistics (Banta, 1987). While the application of statistics had been restricted to a few fields until the 19th century, it gradually extended to other areas since then including physics, astronomy, and recently, air quality (Porter, 1995). At their core, statistical models aim at approximating a relationship between dependent and independent variables, with regression being the most commonly used method, a term that was coined by British statistician Francis Galton back in 1885 when he studied the trend of heights within families (Benirschke, 2004; Galton, 1888; Galton, 1889). The method however precedes the name, with the use of regression starting years before the term was introduced, dating back to the beginning of the 19th century with linear regression being applied to questions in astronomy, such as determining orbits of comets, while the least squares method attributed to Adrien-Marie Legendre and Carl Friedrich Gauss was developed in the early 1800s (Agarwal et al., 2014; Stephen, 1981). In the start of the 20th century, some statisticians introduced the idea of non-linear regression, trying to explain more complex systems (Fisher, 1922). Since then, as computational capacity increased dramatically in the past few decades, regression analysis has been widely used in most scientific fields."

References

Agarwal, R., & Sen, S. (2014). *Creators of Mathematical and Computational Sciences*.

Banta, J. E. (1987). Sir William Petty: Modern epidemiologist (1623–1687). *Journal of Community Health, 12*(2-3), 185-198.

Benirschke, K. (2004). Francis Galton: Pioneer of Heredity and Biometry. *Journal of Heredity, 95*(3), 273-273.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222*(594-604), 309-368.

Galton, F. (1888). Co-Relations and Their Measurement, Chiefly from Anthropometric Data. *Proceedings of the Royal Society of London, 45*, 135-145.

Galton, F. S. (1889). *Natural inheritance*. London: Macmillan.

Porter, T. M. (1981). Social Interests and Statistical Theory: Statistics in Britain, 1865-1930. *Science, 214*(4522), 784-784.

Porter, T. M. (1995). *Trust in Numbers*: Princeton University Press.

Stephen, M. S. (1981). Gauss and the Invention of Least Squares. *The Annals of Statistics, 9*(3), 465-474.

Reviewer #2:

General Comments

In the manuscript, four observation-based machine learning models are developed to predict the top 30 and the 4th highest maximum daily 8-hour average (MDA8) ozone (O_3) concentrations as a function of emissions, meteorological factors, and large-scale climate patterns in Southern California, USA. The top O_3 concentrations, especially the extreme statistics of O_3 concentration, are very difficult to accurately predict. The results show that these four models can explain most of the variations of the observed high O_3 concentrations. The study has examined the applicability of these built models in the South Coast Air Basin (SoCAB) and provide alternative methods for predicting top O_3 concentrations in other regions. I would recommend publication in Geoscientific Model Development after consideration of the following comments.

Specific comments

1. As the results shown in Figure 2, compared with the observations, all of the four models tend to slightly overestimate the lower MDA8 O3 concentrations and to underpredict the higher ones. The four models have very small mean bias (MB, around 1ppbv) when predicting the top30 MDA8 O3 concentrations (shown in Table S3), but they all have higher MB with the average ~10 ppbv underestimation on the 4th high MDA8 O3 (shown in Table 2). As shown in Figure 3, more than 90% predicted O3 concentrations are lower than the observations, which is consistent to the underestimations on the higher MDA8 O3 shown in Figure 2. It indicates that the relationships between model inputs and predicted ozone are different at different ozone levels even addressing the highest 30 MDA8 O3 concentrations. I wonder whether lower MB and RMSE for predicting the 4th high MDA8 O3 would be expected with the empirical models developed using much higher MDA8 O3 (for example, using the data on the top 15 MDA8 O3 days).

    We thank the reviewer for pointing this out. We trained all these four models with the top 15 MDA8 ozone and made the 4^{th} highest MDA8 ozone predictions using these four models, and the results shown below. The model performances for the top 15 MDA8 ozone and the 4^{th} highest ozone predictions are improved, especially the traditional regression models (GAM and MARS). The reason we did not use the top 15 MDA8 ozone days to build the models in this study is that we wanted to include more potential meteorological factors that have an impact on peak ozone formation and make the models more robust with a relatively large amount of data to avoid type II errors. The RH at 850 mb and ENSO index were insignificant when we used the top 15 MDA8 ozone days to build the models. We put this table in the supplement information (Table. S5).

Table S5. Summary of statistical results of the top 15 MDA8 ozone concentrations and the 4^{th} highest ozone predictions using four methods at Crestline site.

Method	**Top 15 MDA8 ozone days**			**4^{th} highest MDA8 ozone**		
	Mean Bias (ppbV)	**R^2**	**RMSE (ppbV)**	**Mean Bias (ppbV)**	**R^2**	**RMSE (ppbV)**
GAM	0.02	0.90	8.30	-3.94	0.98	5.64
MARS	-0.27	0.89	8.55	-4.84	0.97	6.76
RF[1]	-0.40	0.85	10.2	-6.09	0.97	8.12

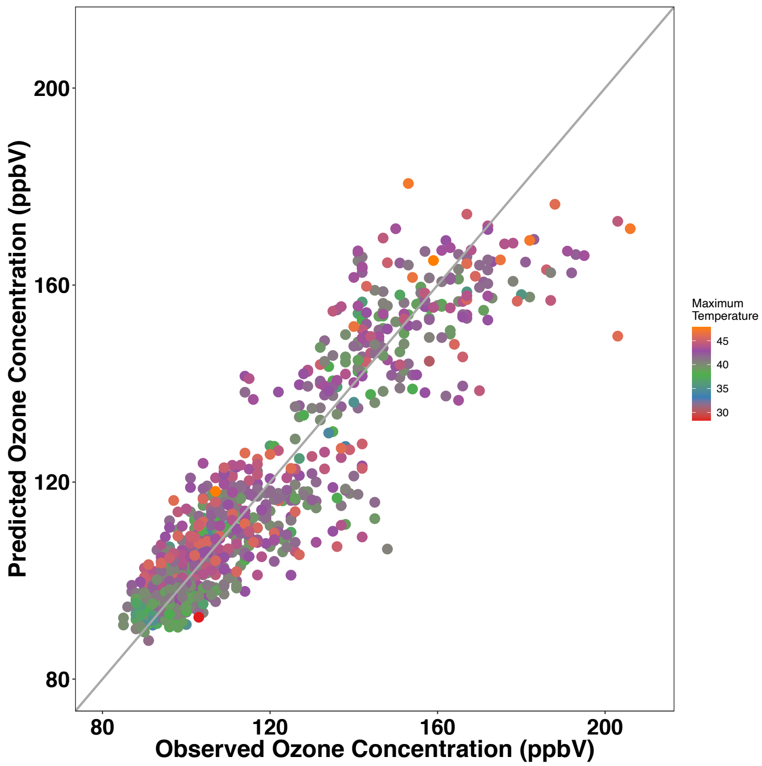
RF²	-0.24	0.85	10.1	-5.89	0.96	8.39
SVR¹	-1.22	0.86	9.92	-4.31	0.93	7.37
SVR²	-1.16	0.88	9.19	-4.60	0.90	9.73

The subscript 1 and 2 in this table: RF/ SVR model with the same variables as GAM model and RF/ SVR model with the optimal combination of the indicators.

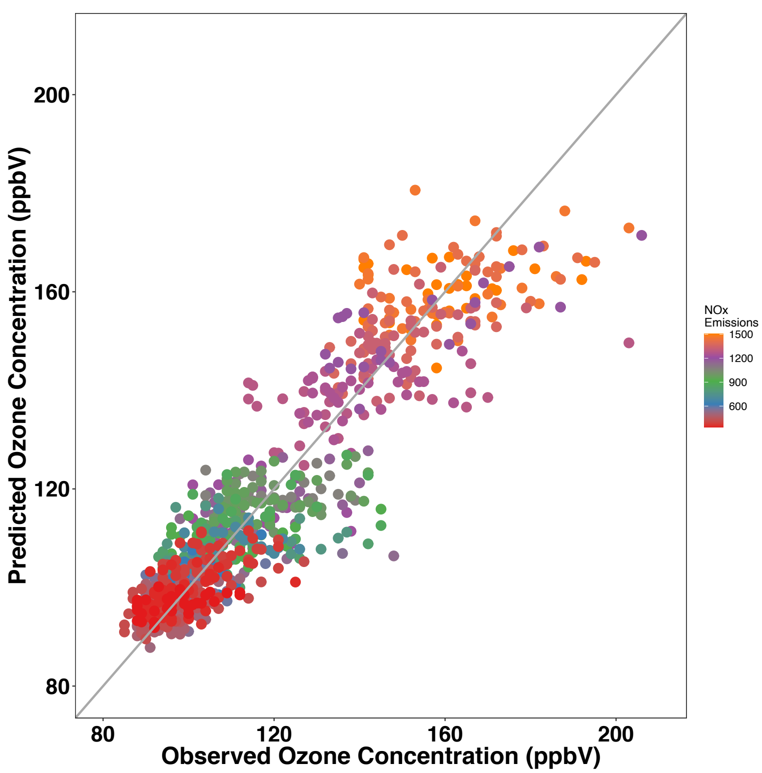
2. As discussed in the Section 3.3 (Limitations), the precursors' emissions in SoCAB and local meteorological variables have been included in the development of the four models. The structure of the built model equations in the manuscript would be applicable for those regions where top MDA8 O3 concentrations are mainly affected by local emissions. However, for the regions where the top MDA8 O3 are significantly influenced by cross-regional O3 transport, more variables might be considered in developing the predicting models (such as the precursors' emissions in surrounding regions).
   Thank you for this suggestion. The average wind speed, wind direction, and the precursors' emissions from upwind areas can be used in the models to explain the air pollutants from upwind areas that influence the smog formation in the downwind areas. In this study, the upwind area of the Crestline site is Los Angeles. The precursors' emissions in Los Angeles are included in the annual NOx and ROG emissions in the South Coast Air Basin. Modified section 3.3 by revising and adding the following sentence between line 490 and line 493.

   "Given that Crestline is downwind of Los Angeles, which then is bordered by the Pacific Ocean, the models using SoCAB emissions capture the upwind conditions. In other regions, such models could be expanded to include both local emissions and upwind states' emissions."

3. In the study, the precursors' emissions have been proved to be the most significant factors impacting the peak O3 levels in SoCAB, and maximum temperature is of relatively high importance among all the meteorological variables. The annual NOx and VOCs emission amounts and maximum temperature from 1990 to 2019 are suggested to be illustrated together with the corresponding 4th high MDA8 O3 (or the top30 MDA8 O3 concentrations) in the Supplementary Information.
   We thank the reviewer for this suggestion. We added the figures below in the supplement information (Fig. S6). The peak ozone concentrations are usually related to a relatively high daily maximum temperature and high precursors' emissions, although some peak ozone days show the opposite relationship. The function of ozone formation is nonlinear with the ambient NOx and VOC in the presence of sunlight. The maximum temperature is highly related to sunlight, so the relationship between maximum temperature and peak ozone levels is nonlinear.



Predicted Ozone Concentration (ppbV)
Observed Ozone Concentration (ppbV)
80
120
160
200
Maximum Temperature
45
40
35
30



Predicted Ozone Concentration (ppbV)
Observed Ozone Concentration (ppbV)
80
120
160
200
NOx Emissions
1500
1200
900
600

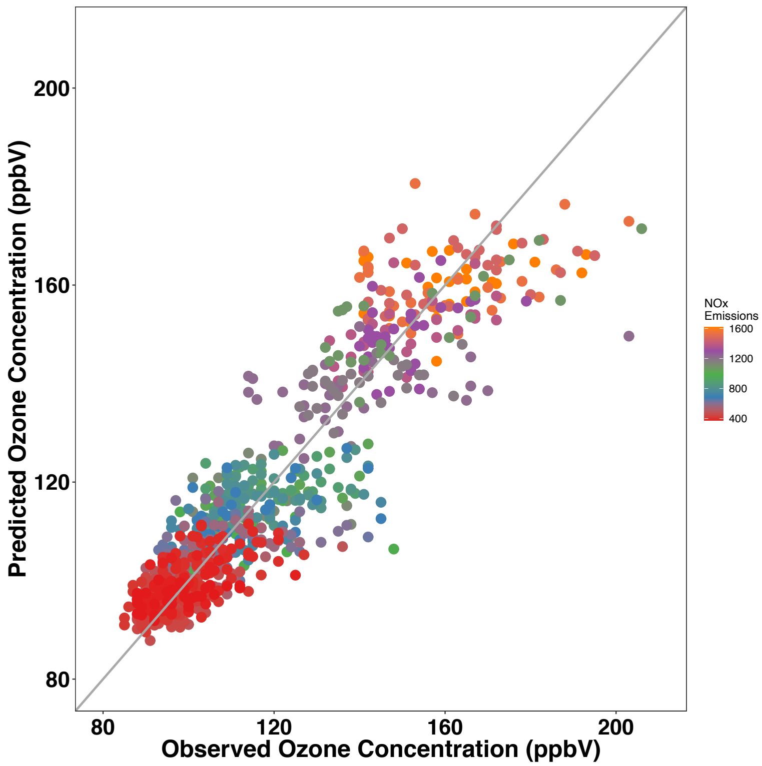


Figure S6. Observed and predicted top 30 MDA8 ozone concentrations with the corresponding annual NOx and VOC emissions and maximum temperature from 1990 to 2019 at Crestline site (the color of the points shows the value of maximum temperature, annual NOx and VOC emissions).