Authors' Responses (in red) to the Comments on acp-2019-642 by Anonymous Referee # 1 and Changes Made (in blue) in the Manuscript

This manuscript separated short-term synoptic-scale fluctuations from long-term baseline component embedded in the daily maximum 8-hr ozone time series using a filter and estimated the limit of air quality model's accuracy (or predictability/uncertainties of air quality prediction). This is an interesting topic for air quality prediction.

The authors thank the referee for recognizing the importance of our work.

But to my surprise, the authors did not even consider lead time when discussing air quality predictability (or limit of air quality prediction). What is the configuration of the air quality prediction? Was this one-day prediction? Two-day prediction? Prediction uncertainties/errors will change significantly with different lead time.

It seems that the referee has misunderstood the modeling simulations we have examined in this paper. Please note that our paper focused on the evaluation of retrospective simulations of 21-years (1990 to 2000) of the daily maximum 8-hour average ozone concentrations over the contiguous United States (CONUS) with the fully-coupled Weather Research Forecasting (WRF) meteorological model and the Community Multiscale Air Quality (CMAQ) chemical transport model. While the concepts presented in our paper are applicable to examining air quality forecasting products, the results of our study reflect the prediction capability of the model based on retrospective simulations and **not** air quality forecasting. To ensure better characterization of the prevailing meteorology (i.e., synoptic forcing) for these retrospective 21-year simulations, four-dimensional data assimilation (FDDA) was utilized following the methodology suggested by Gilliam et al. (see *Atmos. Environ*., Vol.53, pp 186-201, 2012) and modified for fully-coupled meteorology-chemistry model applications as described in Hogrefe et al. (see *Atmos. Environ.*, Vol. 115, pp 683-694, 2015). The modeling set-up and performance evaluation of these historical multiyear WRF-CMAQ simulations have been published by Xing et al. (2015), Gan et al. (2015), and Astitha et al. (2017) as referenced in our manuscript. The following material has been added in the revised manuscript.

Expanded the model description in Section 2 (added after page 3 line 7 in the original manuscript) as follows: "To ensure better characterization of the prevailing meteorology (i.e., synoptic forcing) in the retrospective 21-year WRF-CMAQ simulations, four-dimensional data assimilation (FDDA) was utilized following the methodology suggested by Gilliam et al. (2012) and modified for fully-coupled meteorology-chemistry model applications as described in Hogrefe et al. (2015). The model set-up and performance evaluation of these historical multiyear WRF-CMAQ simulations have been published by Xing et al. (2015), Gan et al. (2015), and Astitha et al. (2017)."

It is also surprising to see the authors suggesting improving simulation of the baseline concentration by focusing on the quality of the emission inventory and the model's treatment for the slow-changing atmospheric processes. I have no question for improving emission inventory, but I am confused by improving the slow-changing atmospheric processes.

A number of papers have been published, documenting the importance of the baseline (longer-term) forcing embedded in ambient ozone data (see e.g., Rao et al., 1996 and 1997; Hogrefe et al., 2000; Rao et al., 2011; Porter et al., 2017; Astitha et al., 2017; Luo et al., 2019). As noted in Astitha et al. (2017), the baseline level and the strength of the synoptic forcing are to be viewed as the necessary and sufficient conditions for observing peak ozone levels. Rao et al. (2011), Hogrefe et al. (2000), Astitha et al. (2017), Porter et al (2017), and Luo et al. (2019) have demonstrated that when the magnitude of the baseline level is low, there will not be ozone exceedances of the USA's National Ambient Air Quality Standard no matter how strong the synoptic forcing is. In this study, we are working with the model (WRF-CMAQ)-predicted and corresponding observed time series of the daily maximum 8-hour ozone concentrations during 1990 to 2000 at various monitoring locations over CONUS; therefore, the Nyquist interval here is 2-days. Using both Empirical Mode Decomposition (EMD) and KZ filtering, we separated the synoptic forcing (time scale < 24 days) and baseline (time scale > 1 month) forcing embedded in the time series of observed and modeled daily maximum 8-hour ozone concentrations. To illustrate, the results of the application of EMD to the daily maximum 8-hr ozone time series data measured at Altoona, PA are presented in Fig. 1 below. The top left panel displays the raw ozone time series while the top of the right panel shows its power spectrum. The 7 intrinsic mode functions (IMFs) and the residual on the left side, and their corresponding power spectra on the right reveal that most of the synoptic-scale features in ozone data are reflected in IMFs 1 and 2. The baseline ozone is extracted by removing the first two IMFs from the raw ozone time series.

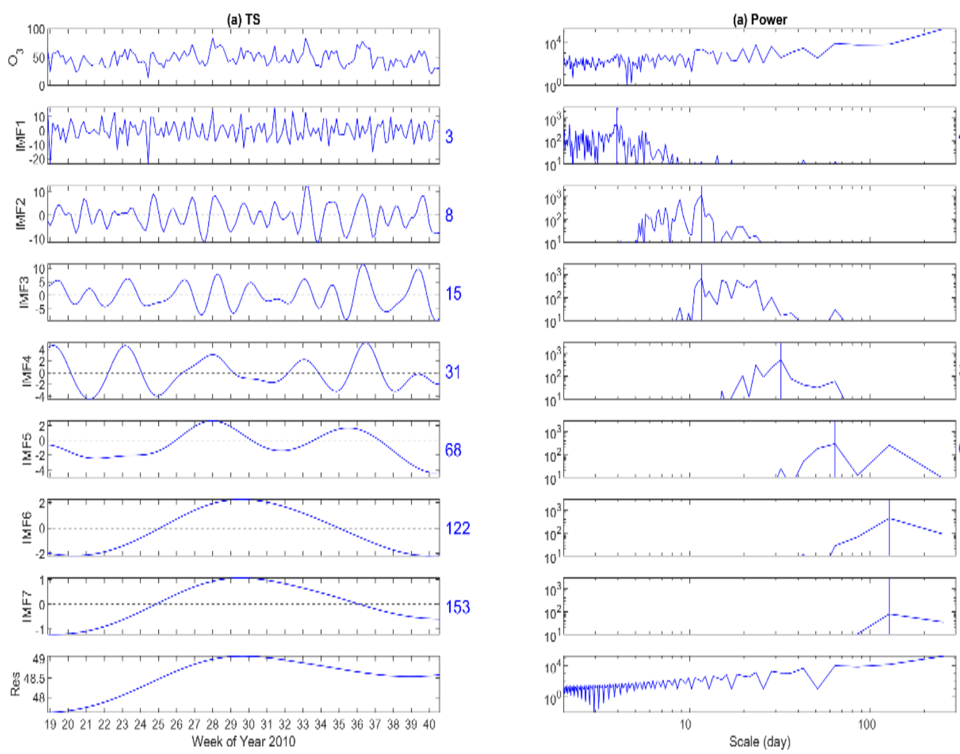

Figure 1. Results of the application of the EMD technique, which is designed for analyzing non-stationary and non-linear time series data, to the daily maximum 8-hour ozone time series data at the Altoona, PA site. The numbers on the right side represent the time scale (in days) associated with each IMF. Note, the power spectrum of raw ozone time series shows that the energy in the 1-10 days (SY forcing) is an order of magnitude less than that in the longer (baseline) time scale.

The scale separation achieved from the application of EMD and KZ filter, displayed in Fig. 2, reveal that the results are quite similar. It should be noted that the short-term component (SY) extracted from the KZ filter as well as EMD's high-frequency IMFs 1 and 2 resemble white noise process.

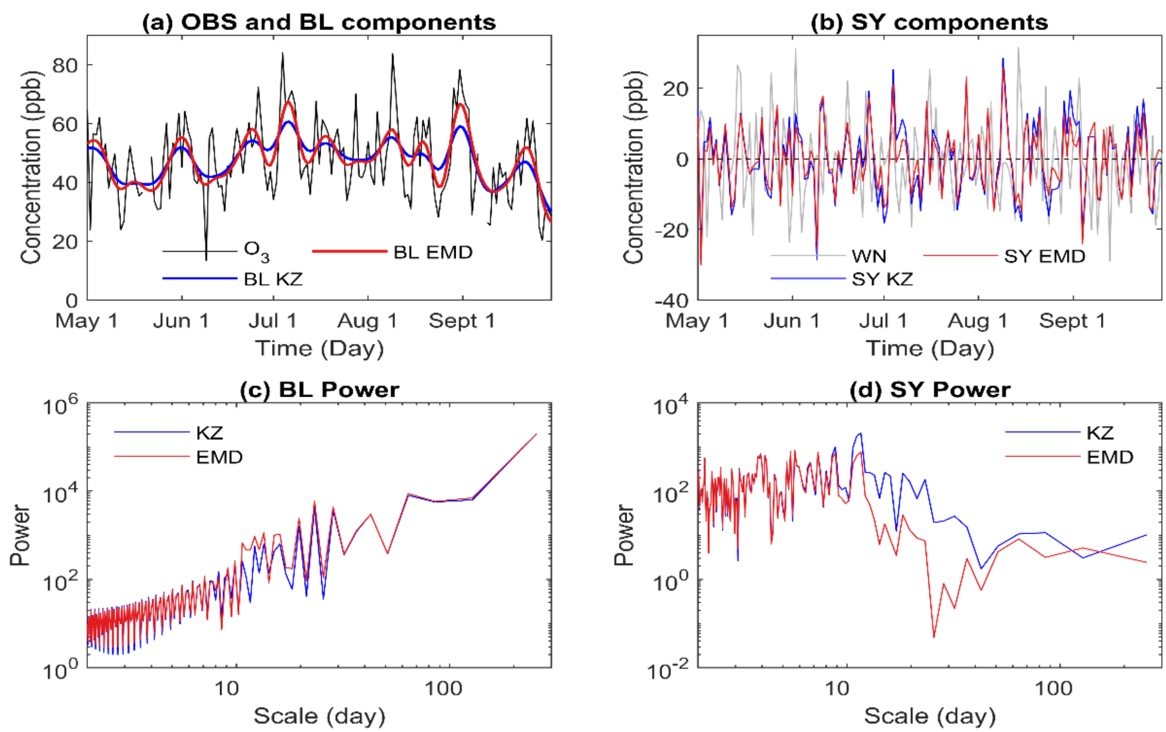

Figure 2. The top left panel displays the raw daily maximum 8-hour ozone time series together with the baselines extracted from the KZ filter and EMD while the top right panel reveals the similarity between white noise and synoptic forcing imbedded in these observed ozone time series. The bottom two panels compare the power spectra of the baseline forcing (left) and the synoptic forcing (right) derived from KZ filtering and EMD (sum of IMF 1 and IMF2). Notice that most of the energy in the baseline time series is in the longer time scale while most of the energy of the short-term component is in the high-frequency range. The similarity of results from both scale separation techniques demonstrates that the two scales of interest (i.e., baseline and synoptic forcing) have been extracted reasonably well.

This material to demonstrate good scale separation has been added in the revised manuscript.

Added new material to the beginning of Section 3.1 (page 4 line 15 in the original manuscript): "Using both Improved CEEMDAN and KZ filtering, we separated the synoptic forcing (time scale < 24 days) and baseline (time scale > 1 month) forcing embedded in the time series of observed and modeled daily maximum 8-hour ozone concentrations. To illustrate, the results from the application of Improved CEEMDAN to the daily maximum 8-hr ozone time series data measured at Altoona, PA are presented in

Fig. 1. The top left panel displays the raw ozone time series while the top of the right panel shows its power spectrum. The 7 intrinsic mode functions (IMFs) and the residual on the left side, and their corresponding power spectra on the right reveal that most of the synoptic-scale features in ozone data are imbedded in IMFs 1 and 2. The baseline ozone is extracted by removing the first two IMFs from the raw ozone time series. To illustrate the concept of the ozone baseline, DM8HR time series measured in 2010 at Altoona, PA is presented in Fig. 2a together with the embedded baseline concentration as extracted by the KZ5,5 and Improved CEEMDAN. It is evident that high ozone levels are always associated with the elevated baseline. The difference between the raw ozone time series and baseline, denoted as the short-term or synoptic forcing (SY), is displayed in Fig. 2b. The power spectra, displayed in Figs. 2c and d, reveal both methods yielded good scale separation. Due to the good agreement between both scale separation techniques, only the results from the KZ filter are presented for the remainder of the manuscript."

The huge advances of weather prediction during the past few decades has been focusing on 1-day or 2-day prediction. On such short-term synoptic-scale weather processes, our weather prediction did excellent job and has been improved through years. Such improvement can benefit air quality prediction significantly (Zhang et al., 2007). I would thus imagine that such short-term practical predictability of air quality can be much improved through better model treatments and better initial conditions of meteorological and chemical variables, as well as emissions.

As stated before, our study deals with evaluating the retrospective air quality simulations, not modeling results from an air quality forecasting effort. Air quality modeling uncertainty even for the retrospective modeling cases, outside of the chemistry formulation and boundary conditions, is attributed primarily to meteorology and emissions inputs. Vautard et al. (*Atm. Environ*., Vol. 53, pp 15-37, 2012) concluded that major challenges still remain in the simulation of prevailing meteorology (e.g., errors in wind speed, PBL, nighttime meteorology, clouds) in retrospective air quality modeling. Based on retrospective ozone episodic modeling with the WRF-CMAQ model using various sets of equally likely initial conditions for meteorology along with FDDA, Gilliam et al. (2015) confirmed the presence of sizable spread in WRF solutions, including common weather variables of temperature, wind, boundary layer depth, clouds, and radiation, thereby causing a relatively large range of ozone concentrations. Also, pollutant transport is altered by hundreds of kilometers over several days. Ozone concentrations of the ensemble varied as much as 10–20 ppb (or 20–30%) in areas that typically have higher pollution levels.

We acknowledge that air quality forecasting often is concerned with accurately predicting transient pollution events. The analysis of modeling uncertainty in our retrospective simulations that employed data assimilation suggest that the largest improvement in forecast accuracy can be achieved through implementing bias correction techniques (i.e. by correcting for systematic errors such as those caused by emission uncertainties that manifest themselves primarily in the baseline) as demonstrated by Kang et al. (2008 *JGR-Atmospheres*, Vol. 113, Issue D23308; 2010 *Atmos. Environ*., Vol. 44, 18, pp 2203-2212). However, we do not dispute the fact that any advances in predicting short-term atmospheric processes or phenomena, be it through model improvements, better initial conditions, or ensemble techniques, may help to reduce the portion of the prediction error that is not due to systematic biases. Rather, we argue through our analysis of both observations and retrospective air quality simulations that, at least for retrospective air quality planning applications, the focus of model development and evaluation efforts should be on longer time scales. The following discussion has been included in the revised manuscript.

Expanded the relevant sentences in the conclusions (page 7, lines 9 – 11 of the original manuscript) as follows: “To improve regional-scale ozone air quality models, attention should be paid to accurately simulate the baseline concentration by focusing on the quality of the emission inventory and the model’s treatment for the boundary conditions and slow-changing (operating on sub-seasonal, seasonal, and longer-term time scales) atmospheric processes. Also, errors in reproducing the synoptic forcing can possibly be reduced with high-resolution meteorological modeling using appropriate data assimilation techniques.”

Other specific comments: Many of the concept/discussion regarding inherent/practical predictability, reducible/irreducible uncertainties are questionable/wrong, or different from those used in weather prediction. For example, emissions are definitely reducible uncertainties and factors in practical predictability. Please carefully define those terms/concepts and refer to normally used/accepted definitions. The writing is overly concise, particularly in many cases where detailed explanation is needed. References: Zhang, F., Bei, N., Nielsen-Gammon, J. W., Li, G., Zhang, R., Stuart, A., & Aksoy, A. (2007). Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. Journal of Geophysical Research, 112, D04304. https://doi.org/10.1029/2006JD007429 Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-642, 2019.

It is difficult to identify exactly which terms the reviewer is referring to in comment that terms are not defined or are not defined as “normally used/accepted”. From our perspective, we are consistent with the reviewer’s example (we agree that uncertainty in retrospectively simulated ozone concentrations can be improved with improvement in emissions characterizations). The confusion in the definition of terms may be more to do with the reviewer’s misunderstanding that the manuscript was focused on air quality forecasting rather than on analysis of retrospective simulations. We agree that terminologies between the weather forecasting community and the air quality modeling community may differ, but we believe that most air quality modelers are familiar with the terminology used in this paper. We respectfully request the reviewer to read the articles by Rao et al. (January 2011 issue of the *Bull. Amer. Meteor. Soc.*, pp 23-30), Dennis et al. (2010), Solazzo and Galmarini (2015), and Gilliam et al. (2015) included in the reference list.