Reply to Referee #2:

We thank the reviewer for the careful reading of the manuscript and helpful comments. We have revised the manuscript following their suggestions as is described below.

(1) What is the definition of "ensemble spread?" Is it a standard deviation, a peak difference, a peak-to-peak difference? The reported ozone spreads, for example, seem small compared to the spread evident on the plots.

The definition of "ensemble spread" has been added in the third paragraph of section 2: "*The ensemble spread is the average difference between the individual ensemble forecasts of a quantity and the ensemble mean forecast of the quantity.*" The reported ozone spreads are the ensemble spread. They are small compared to the spread evident on the plots, which is the maximum difference between two extreme members.

(2) All the plots need to be increased in size to be readable.

All the plots have been updated with the bigger size.

(3) In section 4.1, line 20: It is better not to use altitude words to describe chemical concentrations. Although "elevate", "rise", and such are commonly used, they can lead to confusion. "Increase" would be clearer when one is talking about an increase in concentration rather than the movement of an air mass.

In section 4.1, line 20, the word "elevate" and "rise" have been changed to "increase".

(4) Section 5 is potentially very interesting, but the authors do not say enough about it. They wish to conclude that the relatively small initial condition perturbations shown in the previous section are more important than the choice of PBL scheme. Some more detail would help to convince the reader. For example, what are the primary differences in concept between the PBL schemes? Is any observed difference in performance due to the PBL scheme, to its coupled surface layer scheme, or to differences in the surface fluxes induced by those schemes? Are the BL heights calculated by the different schemes truly directly comparable?

In section 5, the primary differences in concept between the PBL schemes have been included and changed: "*The planetary boundary layer (PBL) is responsible for vertical sub-grid-scale fluxes due to eddy transports in the whole atmospheric column, not just the boundary layer. The PBL schemes determine the flux profiles within the well-mixed boundary layer and the stable layer, and thus provide atmospheric tendencies of temperature, moisture (including clouds), and horizontal momentum in the entire atmospheric column. WRF ARW v2.2.1 includes MRF, YSU, and MYJ schemes. The MRF scheme employs a so-called counter-gradient flux for heat and moisture in unstable conditions. The PBL height is determined from a critical bulk Richardson number. The YSU scheme is the next generation of the MRF scheme, which also uses the counter-gradient terms to represent fluxes due to non-local gradients. The PBL height is defined* *from buoyancy profile. MYJ scheme represents a nonsingular implementation of the Mellor-Yamada Level 2.5 turbulence closure model (Mellor and Yamada, 1982) through the full range of atmospheric turbulent regimes. PBL height is derived from the prognostic turbulent kinetic energy (TKE) equation*.", and "*In order to investigate the impact of different PBL parameterization schemes on the ozone simulation, we have conducted ensemble forecasts using two other PBL schemes, the YSU PBL scheme and the MRF PBL scheme coupled with Monin-Obukhov surface layer scheme (Monin and Obukhov, 1954) in addition to the MYJ PBL scheme coupled with MYJ surface layer scheme (Janjic, 1996; 2002) used in the control case. All the PBL schemes are coupled with Noah land surface model (Chen and Dudhia, 2001).* " Fig. 11 has been changed to the ensemble mean with different PBL schemes, which can clearly show the ozone forecast differences caused by different PBL schemes. Fig. 12 has also been changed to the ensemble means with different PBL schemes, which show differences in meteorological fields due to the PBL schemes. According to the concepts of different PBL schemes, the BL heights calculated by the different schemes are truly directly comparable.

(5) Section 6 could also be expanded. I am not sure I agree that the time evolution "basically agrees well with the observations." The three days are quite different. Why is the predictability different on the different days? The provided explanation is not very convincing, but the authors probably have enough information to provide a better one.

Section 6 has been expanded. We have included one more figure (Fig. 14) to show the different meteorological predictability in all selected days (four days), which basically interpret the different ozone predictabilities during these days. The reason for different predictability on different day has been explained and included in Section 6: "*The practical predictability of the atmosphere is dependent on model errors and initial conditions uncertainties. In our case, since we have used the same model and set up for all selected days, the difference in predictability can only be caused by initial condition errors. We interpolated NCEP GFS-FNL reanalysis data to our model domain to produce initial conditions for every day. The initial condition error on different day is apparently different. Furthermore, the error growth in the model is also dependent on the flow regime, which means even the same initial error grows differently under different flow regime (such as Zhang et al. 2007).*"

(6) The Conclusions should be checked for consistency. Are the uncertainties in ozone primarily due to initial conditions, as stated on the paragraph beginning on line 8? Or are model and emissions uncertainties more important, as implied by the second paragraph? Perhaps a clearer distinction between random and systematic errors would help to clarify. The last paragraph makes an important point, that improved initial conditions may not be possible, and therefore a prediction system must be robust against expected uncertainties. This is certainly true for small random errors in the analysis, as shown here, but may not be true for systematic errors, which might be more subject to improvement.

In the present paper, we didn't compare the uncertainties in the ozone prediction caused by different sources (such as meteorological initial condition error, model errors, or emission sources). Our analysis focus on meteorological initial condition error and PBL schemes (part of meteorological model error). Systematic errors can also be reduced through using ensemble forecasts using different models or different schemes, which is another topic. The 4th paragraph in Section 7 has been changed as reviewer suggested: "*The uncertainties in the ozone prediction, especially during the peak ozone period, due to the meteorological initial condition error, are dramatic. The magnitude of the ensemble spread varies with the adapted PBL scheme in the model, which can affect the PBL height, wind, and temperature and in turn affect the* O_3 *simulation results. The differences in ozone simulations caused by different PBL schemes mainly occur during nighttime, early morning hours, and peak ozone hours. This kind of differences can reach 5-10ppb. In addition, the ensemble spread of surface ozone also varies with different meteorological episodes.*"