

## ***Interactive comment on “Application of WRF/Chem version 3.4.1 over North America under the AQMEII Phase 2: evaluation of 2010 application and responses of air quality and meteorology–chemistry interactions to changes in emissions and meteorology from 2006 to 2010” by K. Yahya et al.***

**Anonymous Referee #1**

Received and published: 12 March 2015

A comprehensive model evaluation study on the WRF/Chem performance for simulating meteorology and air quality over two years with two different configurations (offline and online), respectively, is presented. Effort has been made in putting all the analysis together and trying to make meaningful presentations of the data. It is very challenging to perform mechanistic evaluation of air quality models over different years with

C152

so many uncertainties in meteorology, emissions, and ICONs/BCONS. It often entails more advanced skills and techniques to draw credible conclusions about a model's responses to some specific changes over the years by eliminating or reducing interference from other uncertain factors. However, the authors are trying to achieve the goal by simply comparing the model results with observations using the simple statistics (Corr, NMB and NME) and some plots. As the authors pointed out that the main objectives of the Part II paper are to examine whether the model has the ability to consistently reproduce observations for two separate years, as well as to examine whether the trends in air quality and meteorology-chemistry interactions are consistent for both years. But after reading the manuscript from the beginning to the end, the answers to the above questions are not there. Throughout the manuscript, the authors were talking about statistics superficially without in-depth analysis about what caused the agreement/disagreement. When pairing cell-averaged model predictions with point measurement data in space and time (incommensurability), how much confidence do you have in terms of the good/bad performance of a model for different years with a few percentage difference in NME? I don't oppose using the statistics to perform model evaluations, but it seems too much for me if the analysis is heavily dependent on these numbers and the conclusions were drawn based mainly on these numbers.

Comparing the diurnal variations (Figure 4) using the whole year and all site data doesn't make sense to me. Considering all the averaging effect through space and time, to relate temperature with O<sub>3</sub> concentrations in this context is very weak.

In Section 4, I expected to see some in-depth analysis about the model's response to the changes in emissions and meteorology and this should be the central point the authors are trying to make in this manuscript. But after I read the entire section, I was disappointed, because it simply listed the increase or decrease of the species from one year to another with very basic speculations (and some of them are known facts) and the connection between model response and input changes simply wasn't made. The model's response should be reflected (for example) under the percentage changes in

C153

emissions, under the similar weather conditions, does the model respond to the same percentage changes in pollutant levels as it was revealed in the observations.

---

Interactive comment on Geosci. Model Dev. Discuss., 8, 1639, 2015.