This paper describes the performance of a chemical transport model (CTM) in simulating the observed meteorological and air quality parameters in Mexico City during a recent field campaign. While it is useful to obtain statistics regarding model performance, this study is similar in many respects to previous studies that have quantified the performance of CTM predictions in urban areas using operational measurements. The authors have not articulated what new findings were obtained from the evaluation presented in this study. More insights into why predictions of ozone and other trace gases were higher or lower than observed could have been achieved had the authors utilized more of the wide range of surface and aircraft measurements collected during MILAGRO.
Major comments:

1) The methodology employed by this study is similar to many other meteorological and air quality modeling studies. The authors do not put their model results into context of previous CTM simulations of Mexico City or those obtained for other urban areas. Therefore, it is not clear how the quantification of model performance can be used by others or for what purpose.

2) A major problem in this paper is that it does not utilize other surface and aircraft data to evaluate and interpret the model predictions. There are numerous trace gas measurements made at other surface sites and from several research aircraft. At the end of the paper, the authors state that additional comparisons are underway using those data, but those comparisons are needed in this study. To predict ozone correctly for the right reasons requires not only a reasonable emission inventory and a good prediction of the meteorology, but also a good prediction of NOx and VOCs in the region. NOx is evaluated to some extent, but VOCs are totally ignored.

3) Because the field campaign data are largely ignored, this type of modeling study could have been done for any period and for many large cities with an air-quality monitoring network. While the terrain in Mexico City is a challenge for CTMs, there is little evaluation on how the model simulates the thermally-driven circulations other than a direct comparison of observed and simulated surface winds.

4) A discussion of the impact of the chosen model configuration on the predicted air quality parameters is needed. A limited domain is chosen with fixed boundary conditions; therefore, multiday variations in the background ozone and CO mixing ratios would affect the local results to some extent. This could have been quantified given the available data during the field campaign. Emissions outside of the MCMA also seem to be ignored. So sources from other large cities in central Mexico are ignored as well that could contribute to background values observed in Mexico City.

Other Comments:
Much of this material is more appropriate for the WRF Users guide than in this study. It lists options in WRF, that are not even used in this study and it is not apparent on how this discussion contributes to the paper.

The authors now describe which options are used in this study, but provide no reason why these are used as opposed to others available in WRF/Chem.

No discussion is included on the emissions employed for areas outside of Mexico City. While emissions from Mexico City are the dominant factor contributing to photochemical production of ozone in the area, there are several large nearby cites that could contribute to regional background values of ozone and other trace gases. Were emissions outside of MCMA set to zero?

The authors need to include a brief description or table on the initial and boundary conditions for trace gases. I would expect temporal variations in longer-lived species, such as in ozone and CO, that would affect local mixing ratios on the order of 10-20 ppb - which is not small. Instead of using default values, why not use profiles of trace gases obtained from aircraft measurements or ozonesondes to provide more appropriate boundary conditions that vary from day to day. The authors state that the results were insensitive to choice of initial and boundary conditions, but do not describe how that conclusion was reached.

Please provide more specific references that are available, than just a reference to an entire workshop. There are also numerous papers from a variety of mesoscale models on this topic as well.

I suspect that a 3-km grid spacing is sufficient to resolve most of the circulations influenced by terrain in the region, especially when one is evaluating basin-averaged meteorology, ozone, and CO. But mesoscale model predictions still contain errors in the timing and structure of those circulations at individual stations that
are not necessarily due to urban effects. Although the RAMA stations are located in an urban environment, observed winds at those stations are largely influenced by the thermally-driven flows associated with heating/cooling of the terrain and larger-scale synoptic forcing.

Page 1338, line 23: The authors first mention the problem with the YSU scheme of predicting a boundary layer depth at night that is the height of the lowest model level. This is a well-known problem in the WRF community and a recent paper (Hong et al., Monthly Weather Review, 2006) has described a more up-to-date version that attempts to improve the representation of the stable boundary layer. The text should mention this somewhere. It would be useful for the authors to re-run the simulations with the newer scheme that has been available to the community since June 2008, since the version of the YSU scheme used in this study is out of date.

Page 1339, lines 1-18: The authors should also describe now predicted surface winds (presumably at the lowest model level) are compared to observations (which are most made at a lower height). If the model level is higher than the observation height, then the simulated winds should be higher than observed (unless the simulated winds were extrapolated to the observed height). As stated by the authors, the biggest factor contributing to the over-prediction in wind speed is likely the omission of an urban canopy parameterization. WRF/Chem does have an urban canopy parameterization, but there may not be sufficient information to define building information needed for this parameterization.

Page 1340, line 14: While it is probably correct that the positive bias in ozone is related to NO emissions being too small, the PBL also plays a role. Since the predicted PBL is too low it should have led to ozone lower than observed if the emissions were correct. So, the underestimation in ozone is exacerbated by errors in the PBL depth. There could be other plausible contributing factors as well.

Page 1341, line 15: The authors note two important sources of SO2 that could affect
the under-predictions of SO2; however, those plumes would be transported into the basin periodically. Was SO2 consistently lower than observed? Or was SO2 just too low when the observations indicated large peaks associated with the Tula and volcano plumes? If SO2 was always lower, then that would point to local sources in the city rather than the two point sources. Estimates for SO2 are available for these sources; why not employ them in the model?

Page 1342, line 1-12: There is a recent paper (Stephens et al., ACP, 2008) on weekend/weekday effect in Mexico City that should be cited. There may be some information in that paper useful for interpreting the findings of this paper.

Page 1342, line 18: I do not understand the author’s speculation of SO2 sources outside of Mexico City to explain the observed weekend/weekday variations.

Page 1342, line 25: The authors correctly state the importance of PBL height for predicting trace gas concentrations. There have been numerous studies in the literature that already have performed sensitivity simulations using CTMs (e.g. CMAQ) similar to those presented in this section. Some need to be cited as well as how the present sensitivity compares with previous findings.

Page 1343: lines 23-24: The authors state that the PBL height compares favorably with Shaw et al. (2007), but do not provide any evidence. Why not utilize those data in this study?

Page 1346, line 15: I agree with the author’s statement that more extensive analysis is needed. The authors perform a relatively straight-forward analysis that has been routinely performed by other investigators. Little new insight regarding photochemical processes in the urban area has been achieved. Given the extensive data collected during MILAGRO, there could have been interesting analyses to provide new contributions to science.

Section 5.4: Three periods are examined in more detail to determine whether air quality
predictions are better for some meteorological conditions than others. The time series and statistics indicate similar levels of performance, although a longer time period with multiple days for each type of meteorological conditions would have been useful. I am not sure much is gained by presentation of this material than what is stated in Section 5.4.4. While they conclude that it would be useful to include wet deposition, the lack of a treatment of wet deposition did not seem to make the performance of that period any worse than the other periods.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 1329, 2009.