Atmos. Chem. Phys. Discuss., 14, C9817–C9820, 2014
www.atmos-chem-phys-discuss.net/14/C9817/2014/

Interactive comment on "Evaluation of a regional air quality model using satellite column NO_2: treatment of observation errors and model boundary conditions and emissions" *by* R. Pope et al.

Anonymous Referee #2

Received and published: 5 December 2014

The paper by Pope et al. on the evaluation of the British AQUM air quality model with the help of OMI NO2 measurements is interesting and provides a very good example of how satellite data can be used to improve model performance. As a consequence, the study is rather technical: satellite data handling is discussed in great detail, and approaches for model repair (boundary conditions, emissions, chemistry) are introduced, but there is no obvious scientific discovery, so the progress in this paper is merely methodological. Still, it is an interesting paper that should be published. One distinct merit of the paper is that it shows how satellite data are useful to evaluate and pos-

sibly even improve multiple aspects of a air quality model, e.g. boundary conditions, emissions, and chemistry.

Generally speaking, the application of the averaging kernels is interesting, and documented extensively. In some places, however, the description and explanation of the AK needs improvement. In equations 1-3, y is a scalar, and A and x are vectors and the notation should be adapted accordingly (e.g. in boldface or with a vector sign). A discussion of why the kernel values over Dartmoor are lower than over London is missing. This should be explained. Kernels depend on a range of aspects: albedo, viewing geometry, clouds, and also location. So what do we see when kernels over Dartmoor and London are different? Are albedo, geometry and clouds all similar between those two locations, and do we see the effect of the different NO2 loading between Dartmoor and London? More generally, it is unclear what the key message is for section 2.1 and Figure 1. The authors should reconsider this section and their discussion – what is the take-home message?

On page 21757 it should be clarified what causes the systematic error in the slant column. Table 1 in Boersma et al. [2004] provides this information. Related to this, the systematic error in the slant column should not be included in Eq. (6) as it does not propagate into a tropospheric column error. Any systematic error in the NO2 slant column density will be largely absorbed by the stratospheric assimilation procedure. Subtracting from the (biased) slant column a similarly biased stratospheric slant column results in a small tropospheric residual slant column. See e.g. Belmonte-Rivas et al. [2014].

The calculation of the time-averaged retrieval error is an interesting aspect of this paper. The authors claim that the original, individual (per pixel) retrieval errors can be reduced by 30-90% by averaging over time because random contributions largely cancel. Although this aspect of DOAS retrievals (averaging out the spectral fitting noise contributions) is well known, and has been applied many times in earlier studies, the authors now provide a practical framework to justify and quantify the time-averaged error. However, one problem in the way they present their results, is that what they define now as the systematic error (caused by errors in the AMF), is not a true systematic error in the sense that it is a persistent and always works in the same way. The authors use the notion that the retrieval error due to AMF errors is 40% of the tropospheric column and systematic. But the AMF error changes in time (on day 1 albedo knowledge may be accurate and on day 2 albedo may be off by +0.02), and in space (for pixel 1 the cloud fraction may be underestimated, for pixel 2 it may be overestimated), and in practice at least part of these individual systematic errors may also cancel in a temporal or spatial average. The formulation in Eq. (10) leaves no room for this now, i.e. all AMF error is 40% and always proceeds systematically. The difficulty obviously is in estimating to what extent the systematic error cancels when averaging over time, but validation studies have provided some hints here (individual comparisons may be off by up to 40% but averaged over an ensemble the systematic differences are more on the order of 20%, e.g. Irie, Hains, Schaub-papers).

In the revised manuscript the above concern should be taken into account, and the authors in their discussion of Figure 2 should also indicate the magnitude of the original errors (without subtracting the pure random contribution). There should also be a discussion of why the errors appear to reduce more over continental Europe than over Sea. And why is the reduction stronger in summer than in winter? Related to more samples in summer?

Specific comments:

P21752, L1-2: how can models 'detect' shipping lanes?

P21752, L18: 5-6% overall or per year?

P21754, L2-3: please clarify the statement here as there was not much of an OMI row anomaly in 2006.

P21754, L7-11: the Irie et al. [2008] citation is irrelevant here, since this concerns

validation of another product than used here. The paper by Irie et al. [2012] is the appropriate paper that discussed the validation DOMINO v2.0 data.

P21754, L17-19: suggest to clarify 'the true vertical distribution'.

P21756, L4: 'with trap'?

P21758, L8-10: how can the left-hand side of Eq. (8) become negative? The authors should explain in what sort of situations this happens, and why.

References:

Irie, H., et al. "Quantitative bias estimates for tropospheric NO 2 columns retrieved from SCIAMACHY, OMI, and GOME-2 using a common standard for East Asia." Atmospheric Measurement Techniques 5.10 (2012): 2403-2411.

Belmonte Rivas, M., Veefkind, P., Boersma, F., Levelt, P., Eskes, H., and Gille, J.: Intercomparison of daytime stratospheric NO2 satellite retrievals and model simulations, Atmos. Meas. Tech., 7, 2203-2225, doi:10.5194/amt-7-2203-2014, 2014.

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 21749, 2014.