Response to reviewer comments on
"Hindcast experiments of tropospheric composition during the summer 2010 fires over western Russia"
by V. Huijnen et al.

First of all, we thank the reviewer for his constructive comments.

The manuscript by V. Huijnen et al. is a model investigation of the atmospheric composition during the period of severe wildfires in the Western Russia in July – August 2010. The goal is to improve accuracy of forecasting air pollution from wildfires. Satellite data from MODIS, MOPITT, IASI, OMI, and SCIAMACHI were assimilated by a CTM forecasting model and the authors believe that accuracy of the model improved. The paper operates with the FRP technique to estimate emissions, as well as with current satellite retrievals. An important drawback is ignoring results of published ground-based measurements. Also a critical analysis of FRP technique has not been presented. However, the paper may be published after revision.

General remarks.
First of all, the title made me to open a dictionary: the word "hindcast" is a modelers' slang. Moreover, there is no explanation for this term in the introduction. Please, add a couple of sentences about hindcast, forecast, reanalysis, etc. Also the word "experiments" in the title is misleading. This is a minor problem, however, if both are explained in the very beginning of the paper.

We chose to use the word 'hindcast' in the title on purpose, as it reflects the meteorological approach of our study. We assume the reader understands the concept of forecast, and reanalysis, which is also made clear from the context, and the reference (Hollingsworth et al., 2008). We now include the following comment to explain the definition of a hindcast run:

For this purpose several hindcast (i.e. retrospective forecast) experiments with different model settings have been evaluated.

The paper uses the FRP approach for the bottom-up emission estimate. However, a better accuracy of the FRP technique versus, say, "active fires" technique is not proven yet. E. g., the "active fires" technique used by Fokeeva et al. (2011) demonstrated much closer agreement with a "top-down" estimate for the Russian fires. Comparisons of FRP emission estimates with other estimates is not presented in the paper. I understand, that the "active fires" technique might be too time-consuming and not usable for operational forecasting, but this should be explained clearly.

Different to what the reviewer writes, we do include a comparison of GFASv1 to GFEDv3.1 for this particular case, and explain key differences by the applied predominant soil type map. A detailed evaluation of the GFAS system is beyond the scope of this paper. We refer to Kaiser et al., 2012 and Heil et al., 2010, which includes detailed global comparisons of the GFAS system to GFEDv3.1, and hence serves as a global evaluation of the GFAS system. Kaiser et al., (2012) show general good consistency between GFASv1.0 and GFEDv3.1.

To our understanding Fokeeva et al. (2011) derived bottom-up emission estimates with a combined Active Fire and Burnt Area approach. Nevertheless, details on the methodology to derive their emissions are not given. Also, despite the apparent agreement with a top-down estimate for the current event by Yurganov et al. (2011), a global assessment of their method, compared to other systems (Finn (Wiedinmyer et al. 2011), GFEDv3 (van der Werf et al., 2010)) is missing, which makes it difficult to comment on general performance.

The Fokeeva et al. (2011) results have been compared to the top-down estimate by Yurganov et al. (2011), and indeed showed good consistency. However, also the Yurganov et al., (2011) method seems connected with high uncertainties, considering the simplicity of their method. Furthermore, the Yurganov et al., (2011) and Fokeeva et al. (2011) emissions were not evaluated in a chemistry transport model, different to the Konovalov et al. (2011) emissions, as well as our current emission estimates. This all makes it hard to comment on their method.

A weak point of the paper is insignificant citing of literature and practically lacking information about experimental techniques and retrieval algorithms. Also, to characterize any algorithm or experimental technique one sentence is enough, but it should be done.

The reviewer is correct that references to observations entering the MACC system were missing in some cases. These are now provided systematically in Table 1. Nevertheless, we hesitate to provide too much information on experimental techniques in the manuscript for all the 7 data products, as this goes beyond the scope of this paper and can be found in the references. Relevant data aspects (averaging kernel, vertical sensitivity range) are provided.

MACC is not the only forecast system: GMI is being developed at the NASA/Goddard, others may also exist.

We thank the reviewer to point this out. We have now included a brief overview of existing forecast systems, see also the response to the other reviewer.

Numerous ground-based measurements of the atmospheric composition (not only AERONET data) for this event have been published and must be compared with the model results (see references below).

The primary purpose of the MACC global system is to model the atmospheric composition at background conditions and in the free troposphere. It is not suitable for air quality applications. Therefore in-situ surface observations cannot simply be used to assess model performance, unless when observations are well characterized as representative background values. This may be questionable for most ground-based in-situ observations during this event, being either close to emission sources of the megapolis, or close to local fires.

This made us decide to focus our model evaluation on the free troposphere. We make the scope of our work more clear in the introduction of our revised manuscript. Nevertheless, the reviewer is correct that ground-based observations still yield relevant information. This is especially important because difference between runs with/without

GFAS is mostly visible in the lowest layers, which is hard to observe from satellites, see also below.

The authors noted: “the highest sensitivity of the [satellite] instruments ... [that] provide data used in the assimilation is to the stratosphere”. It is correct. According to Fig. 1, the sensitivity of IASI column to CO in the bottom layer ~400 m thick is just ~0.15. Real change of partial column (or a change in the computed value due to another emission rate) in this layer should be multiplied by 0.15 before adding to the entire retrieved (or modeled and convolved) column. This is what is usually meant by the low sensitivity of TIR IR sensors to surface emissions. This fact diminishes accuracy of any schemes for top-down estimates based on TIR satellite data only (without ground-based validation). That is why using ground-based data would improve the paper a lot, or, at least, indicate the accuracy of computations. As for CO, it would be useful to compare model surface concentrations (Fig. 7) with observed surface concentrations on the same graph. The same can be done for available surface data for other components.

We acknowledge that satellite data of atmospheric trace gases have in most cases little sensitivity to concentrations at the near surface, where emission changes are dominant, as we also show in our manuscript (Fig. 7). In this respect ground-based data is a useful contribution. However, as noted above the focus of this work is on the modeling of concentrations in the free troposphere and for background conditions, rather than air-quality applications, which are focused on the model performance at the surface. Considering the relatively low spatial resolution of the chemistry model, validation against in-situ surface station data can only be done when this station is well characterized and representative for background concentrations. Such ancillary information is mostly missing (e.g., Elansky et al., 2011). Ground-based total-column information, such as AOD, is less sensitive to such local conditions. Also for CO we acknowledge that total column information is helpful additional information. Therefore we now include a comparison of the model against ground-based CO total columns as measured from the Moscow and Zvenigorod stations. It is found that they are well able to distinguish between runs with/without GFAS emissions, which now provide valuable complementary information to the satellite-based observations of CO.

Added sentences:
Additional to the space-based observations, we evaluate the model runs against ground-based CO total column observations based on spectrometers at the Moscow and Zvenigorod stations, as reported in Yurganov et al. (2011). The Zvenigorod observation station is located 53 km west of the Moscow station. For this evaluation the modeled day-time CO profile has been interpolated to the station location and convoluted with the averaging kernel corresponding to the observations. The magnitude of the total column observations from the two stations at the same day are mostly relatively close, except for 6 and 9 August, when the observations at Moscow are about two times larger than the ones at Zvenigorod. The model columns interpolated at the two stations are always very similar. This illustrates that very local events, causing differences in observations cannot be resolved at the current model resolution.
Different to the evaluation against space-based columns, a distinct difference between the runs with/without assimilation is clear, while both runs that apply GFAS emissions are now relatively close. The rather modest increase in run Assim compared to CNT can

be explained by the low sensitivity of IASI near the surface, as illustrated by the averaging kernel in Fig. 1, and because the a-priori profiles do not contain the high surface concentrations for this particular event.
The model versions with GFAS emissions capture the increase in CO columns during the first 10 days of August, but over-estimate concentrations on 29 July, related to the estimated high peat fire emissions close to Moscow on that day (Fig. 3). This is in contrast to AOD model results in Moscow for this day (see Fig. 4) and suggests that this model bias is caused by the lower resolution of the chemistry model compared to the aerosol model in IFS.
Model mean bias and RMSE during the period of the fires is presented in Table 7. The D+0 hindcasts show a significant model improvement when using GFAS, and best performance is obtained by combining GFAS emissions and assimilated CO data. For hindcast days D+1 to D+3 the degradation in performance compared to D+0 is relatively moderate for all runs. Run Assim-GFAS remains best up to D+3, which is again different from the evaluation of ground-based AOD. Especially the fact that the RMSE does not degrade for the runs with GFAS could indicate that the available observations do not fully constrain the model performance. For specific days with large discrepancies between D+0 and D+3 forecasts (e.g. on 7-8 August, not shown) there are unfortunately no observations.

I agree with the second reviewer that the authors care too much about a better “image” of the paper. The value of the paper would just increase if they demonstrate real accuracy and real problems of the model used.

We find these comments difficult to appreciate: We feel that we are open in presenting not only the successes (e.g., CO total column evolution against MOPITT observations) but also the failures (e.g., problems with high ozone concentrations over Kazakhstan, issues with RMSE and spatial correlation in NO2). We make no claims on model performance near the surface and close to (anthropogenic / fire) emission sources, where the system is likely showing problems in its accuracy. As written above, we have now included a statement of the limited scope of the paper (i.e. the free troposphere) in the introduction.

Also they should look in a wider context, e.g., an interesting question is NO2 perturbation due to fires: the temperatures might be very high, formation of NOx might occur. Does it happen or not? Comparison with in situ measurements before and during fires (see references) in Moscow would be highly interesting. A paper in ACP should not be devoted to testing a model only.

The reviewer points to an interesting topic: the amount of NO2 being produced from fire emissions, which depends strongly on the fire type. However, we think an assessment of in-situ NO2 production is beyond the scope of this paper, also considering the uncertainties in the model emissions at its coarse spatial resolution, and corresponding issues with the representativity of in-situ observations of these short-lived components for model evaluation. The current paper is focused on presenting the phenomenon on a sub-continental scale, answering the relative question to what extend data assimilation and NRT fire emission estimates contribute to the accuracy of the forecasted atmospheric composition. In the introduction we have now more clearly motivated our research, and placed this in a wider context.

Specific remarks.

1. CNT is not spelled out.

We think this name is self-explanatory.

2. IASI CO retrievals during Russian fires are found to be more accurate than those by MOPITT and AIRS (Yurganov et al., 2011). There are traces of (previously used?) IASI data (CO averaging kernel in Fig. 1, also p.31864, line 4, p.31864, line 3. I recommend to mention why IASI CO data were not used, if there is no possibility to compare the data for the same event, but different instruments and algorithms.

We refer the reviewer to Table 1, which shows that IASI CO data is indeed used for the hindcast runs that include data assimilation for the initialization of the hindcasts.

3. For the inversion module COPRAFIT a reference (Eremenko et al., 2008) is missing in the reference list (page 31865).

We include this now.

Summary. The study is devoted to a problem that is very important both practically and scientifically: forecasting air composition perturbations due to fires. The authors have done a good job, but they did not devote enough attention to real accuracy of the specific forecasting scheme. Accuracies of both FRP and satellite data are questionable; this point should be elucidated in more detail. A significant revision of the paper is necessary.

The reviewer would like to see a more quantitative assessment of the system, by including ground-based in-situ observations. As explained above, the usability of such observations is limited, considering the scope of our current system. Studies with local air quality models, such as the one described by Konovalov et al. (2011), are more suitable for this. Nevertheless, we now include an evaluation against ground-based CO total columns, which is indeed a useful addition to the current observations. It helps to assess the performance of the GFASv1 emissions and better discriminates model results where assimilation is included, but with/without GFAS emissions.

An evaluation of the accuracy of emission estimates, in this case based on FRP observations, can only be done indirectly, by applying these emissions in chemical transport models and evaluating results against independent observations. The assessment of the accuracy for this particular case is given in considerable detail in our current manuscript, for a more general assessment of the accuracy of emissions compared to GFEDv3.1 the reader is currently referred to Kaiser et al., (2012). A global assessment of GFASv1 emissions as part of the CTM is beyond this study, but will certainly be subject of future model evaluations.

We are fully aware of uncertainties in satellite retrievals but did not discuss them in depth, just to keep the manuscript compact. We did mention accuracy issues when they were found to be crucial (e.g. for the evaluation of tropospheric ozone and formaldehyde) and refer the reader to literature dealing with satellite retrievals that will explain better the retrieval algorithm. We note that by applying averaging kernels to the model concentrations we compensate for possible retrieval biases.

The reviewer is correct that it is useful to add more complete reporting on satellite retrieval accuracy throughout the evaluation section, to assess the significance of the evaluation. We now provide systematic information on the retrieval accuracy for the validation data throughout section 3, when introducing the observational data. Furthermore we have extended the description and uncertainties of the SCIAMACHY and OMI NO2 retrieval products, see also our response to the other reviewer.

References

Elansky, N.,Mokhov, I.,Belikov, I.,Berezina, E.,Elokhov, A.,Ivanov, V.,Pankratova, N.,Postylyakov, O.,Safronov, A.,Skorokhod, A.,Shumskii, R.,Gaseous admixtures in the atmosphere over Moscow during the 2010 summer, Izvestiya Atmospheric and Oceanic Physics 47 (6), doi: 10.1134/S000143381106003X, 2011.

Fokeeva, E., Safronov, A., Rakitin, V., Yurganov, L., Grechko, E.,Shumskii, R.: Investigation of the 2010 July–August fires impact on carbon monoxide atmospheric pollution in Moscow and its outskirts, estimating of emissions. Izvestiya Atmospheric and Oceanic Physics 47(6) doi: 10.1134/S0001433811060041, 2011.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527-554, doi:10.5194/bg-9-527-2012, 2012.

Konovalov, I. B., Beekmann, M., Kuznetsova, I. N., Yurova, A., and Zvyagintsev, A. M.: Atmospheric impacts of the 2010 Russian wildfires: integrating modelling and measurements of an extreme air pollution episode in the Moscow region, Atmos. Chem. Phys., 11, 10031-10056, doi:10.5194/acp-11-10031-2011, 2011.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, doi:10.5194/acp-10-11707-2010, 2010.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4, 625-641, doi:10.5194/gmd-4-625-2011, 2011.

Yurganov, L. N., Rakitin, V., Dzhola, A., August, T., Fokeeva, E., George, M., Gorchakov, G., Grechko, E., Hannon, S., Karpov, A., Ott, L., Semutnikova, E., Shumsky, R., and Strow, L.: Satellite- and ground-based CO total column observations over 2010 Russian fires: accuracy of top-down estimates based on thermal IR satellite data, Atmos. Chem. Phys., 11, 7925-7942, doi:10.5194/acp-11-7925-2011, 2011.