Authors response to comments of the Referee #1 Prof. N. Moussiopoulos

We thank the Referee #1 Prof. N. Moussiopoulos for the interesting and important comments on our manuscript. All the individual comments are addressed below in red.

This manuscript provides a thorough presentation of Enviro-HIRLAM representing one of the first serious development efforts towards implementing a fully online coupled meteorological and chemical weather model. It contains detailed descriptions of methodology selected and implementation followed, including some coverage of less welldefined aspects of online coupling and performance evaluation. The paper is well written and contains a large amount of information. A section on model applications provides additional insight on the extremely important aspects of evaluation and validation.

As the overall assessment of the present referee, the paper successfully describes the remarkable effort that has been devoted to the development of a state-of-the-art online meteorological and chemical weather model. It is adequately referenced and contains detailed explanations of the main physical mechanisms and selected parameterisations. It also highlights some of the more promising aspects of the coupling idea, both in the area of aerosol-radiation treatment and in cloud microphysics.

The only weak point in the manuscript is the rather sketchy discussion of the extent to which the explicit introduction of all effects will lead to improvements in model performance.

Response:

The paper is focusing mostly on the model description and its applications, therefore it was not much space for detailed discussion of specific effects of different model improvements. Some aspects were published in previous papers, some are still in new specific papers to be submitted (e.g. for pollen applications, operational air quality forecasting).

We have extended this part in the revised version.

Section 4 of the manuscript represents of course an honest attempt to summarise what we know on the effect of coupling in model performance for different applications. The authors are encouraged to provide more explicit comments in this respect. This should be combined with a more thorough discussion on how all parameters required in the various process parameterisations could be fine-tuned (for instance, expanding the comments made in the last four lines of the paper).

Response: The Section 4 is extended correspondingly in the revised version.

In the below listed specific comments references are made to specific lines in the text.

1. Methodology and modelling system structure

a. The model coupling implements aerosol impacts on radiation (direct and semidirect effects) and on clouds (first and second indirect effects), l. 110-111. It appears appropriate to include an explicit reference to COST action ES1004 in the framework of which these effects were extensively discussed.

Response: Thanks, agree. The corresponding reference is included.

b. The cloud feedback module includes some rather advanced approximations (l. 287-293); the reader would welcome more remarks on the extent to which this complex cloud model has been validated.

Response:

Abdul-Razzak and Ghan parameterization for aerosol activation has been extensively tested in many online-coupled weather and climate models. However, the STRACO cloud microphysics scheme with parameterizations of aerosol activation, cloud droplets nucleation, sedimentation, evaporation, self-collection, has never been thoroughly evaluated, only with 1D column HIRLAM. These evaluation results are not ready to be published yet, but will be analysed and published on further steps.

c. The present HIRLAM NWP model core is based on the hydrostatic approximation (l. 127) which could be a serious limitation over complex terrain (l. 508) and/or in cases of nesting down to urban areas. A plan for a transition to a new, non-hydrostatic platform (e.g. HARMONIE, l. 135) is mentioned, but more information in this respect would be helpful.

Response:
Yes, we agree about the limitations and write openly about them. The new version under HARMONIE is only under development and only some elements are realised so far, so it is too early to describe it extensively it in more details at this stage.
The non-hydrostatic HARMONIE-AROME model includes only some aerosol effects. The physics included in this version of HARMONIE has recently been detailed by Bengtsson et al. (2017). HARMONIE-AROME is based partly on Meso-NH (Mesoscale Non-Hydrostatic atmospheric model), which is a cloud resolving model that includes state-of-the-art chemistry and aerosol interactions (e.g. Berger et al. 2016). Meso-NH can, however, not be run as a near real time NWP model, which is possible with Enviro-HIRLAM.
Corresponding extension text is included in the revised version.
The following additional references are included:
Bengtsson, L., U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, K. Ivarsson, G. Lenderink, S. Niemelä, K. P. Nielsen, J. Onvlee, L. Rontu, P. Samuelsson, D. Santos Muñoz, A. Subias, S. Tijm, V. Toll, X. Yang, and M. Ødegaard Køltzow, 2017: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. Mon. Wea. Rev. doi:10.1175/MWR-D-16-0417.1, in press.
Berger A., M. Leriche, L. Deguillaume, C. Mari, P. Tulet, D. Gazen and J. Escobar, 2016: Modeling Formation of SOA from Cloud Chemistry with the Meso-NH Model: Sensitivity Studies of Cloud Events Formed at the Puy de Dôme Station. In: Steyn D., Chaumerliac N. (eds) Air Pollution Modeling and its Application XXIV. Springer Proceedings in Complexity. Springer, Cham

d. The atmospheric chemistry modules implement a wide array of new parameterisations and numerical schemes (page 4, l. 141-192). Although these were obviously validated separately, their combined implementation in a coupled model definitely needs further validation. Did the authors take already actions in this direction, and if not, what are their plans?

Response:
Yes, these chemical schemes/solvers were tested and validated as standalone versions (Reference: Shalaby, A., Zakey, A. S., F., Giorgi, and M.M. AbdelWahab "Coupling of Regional Climate Chem Aerosol Model", Ph.D. thesis, Faculty of Science, Cairo University-Egypt, 2012). Six environmental/smog chamber experiments were used to validate the gas-phase schemes and different chemical solvers as box models.
The Tennessee Valley Authority (TVA) and the EPA chamber experiments were used to evaluate the different gas-phase schemes and different chemical solvers. Namely, TVA005 and TVA006 are designed to test the simple system of NOx; TVA068 is designed to test a simple mixture of VOC with very high NOx. EPA069A, EPA073A and EPA150A are used to validate the schemes with low NOx concentration and high VOC concentration.
Also, the same chemistry schemes/solvers are coupled with the Regional Climate Model (Reference: Shalaby, A., Zakey, A. S., Tawfik, A. B., Solmon, F., Giorgi, F., Stordal, F., Sillman, S., Zaveri, R. A., and Steiner, A. L.: Implementation and evaluation of online gas-phase chemistry within a regional

climate model (RegCM-CHEM4), Geosci. Model Dev., 5, 741-760, doi:10.5194/gmd-5-741-2012, 2012.)
However such validations need to be further continued and completed, this is the issue for further analysis.
Corresponding extension text and the above mentioned additional references are included in the revised version.

e. The aerosol dynamics model introduces a very interesting classification of particles depending both on particle size and particle composition per emission source. This could allow, in theory, a separate per-source type treatment of particles throughout the chemical mechanism. But is there such a procedure (with potential applications in source apportionment) really implemented or planned?

Response:
The particles classification with respect to their size and composition is based on aerosols classification in M7 microphysics module. As for total particulate matter the splitting is in species, the procedure follows guidelines and recommendations associated with different emission inventories. For now, there is no procedure for source apportionment in the model or plans to do that in the nearest future.

f. Specific emission models for anthropogenic biomass burning (e.g. wildfires) are included (section 2.5). These are based on satellite or other inventory-estimates of yearly fluxes that are temporally disaggregated using pre-defined temporal profiles. Are the latter site dependent, and which is the origin of the coefficients used by the authors?

Response:
The Finish Meteorological Institute developed the IS4FIRES (http://is4fires.fmi.fi) wildfires emission inventory. The IS4FIRES inventory provides temporal profiles for emissions disaggregation with site dependency. However, the profiles used in Enviro-HIRLAM runs for this paper are site independent (mean) and are the functions of a local time only. They were also provided by FMI for AQMEII-2 (COST Action ES1004) initiative.
Changes in manuscript:
L257: The biomass burning emissions typically show a diurnal cycle variability, and therefore, corresponding coefficients are applied (*Giglio, 2007*).
Added to reference list: Giglio, L., 2007: Characterization of the tropical diurnal fire cycle using VIRS and MODIS observations, Remote Sensing of Environment, 108, 4, pp. 407-421.

g. The model contains several “urbanisation” features (section 2.7), including a subset of previously proposed urban parameterisations (Martilli, Dupont, Masson, Grimmond et al.). This is an interesting and original approach, but there are several concerns on how it is implemented (please see comment 2c below).

Response:
The details of implementations of different urban modules, our own developments and comparisons of different approaches and modules were published in previous papers (Mahura et al., 2005b, 2006a, 2007a, 2008bc; Baklanov et al., 2005, 2008), so we don’t describe them in this paper. The main approach includes an integration of the urban modules into the ISBA (Interaction Soil- Biosphere-Atmosphere) land surface scheme of the NWP / HIRLAM model. The urban modules are activated only on those grid cells of the model domain where the urban fraction is presented.
More explanations and corresponding references on the above papers are included in the revised version.

h. In l. 372-378 some aspects of the so-called locally mass conserving semi-Lagrangian (LMCSL) transport scheme are described. The description emphasizes the approximate mass-conserving properties of the algorithm for 1st neighbour cells, but one could ask whether and how is mass consistency ensured in the larger scale. In l. 389-390 it is stated "[: : :]Enviro-HIRLAM is not formally wind-mass consistent regarding tracer transport". The authors should discuss possible consequences of this failure.

Response:
We have added a sentence to clarify that mass-wind inconsistency is a minor problem. The traditional HIRLAM is (at least in principle) wind-mass consistent. In Enviro-HIRLAM where all moisture fields are transported with the LMCSL scheme, there is no formal consistency, yet, since precipitation is very similar to that in HIRLAM (except for individual convective systems that are chaotic/unpredictable in their nature), the mass-wind inconsistency is small in practice.

Suggested new version of lines 372-378 in the original text (changes in bold)
As the traditional SL scheme, the LMCSL is not **inherently** monotonic or positive definite. Therefore a posteriori iterative locally mass-conserving (ILMC) filter was developed, Sørensen et al. (2013).
……
It should be noted that the dynamical core in Enviro-HIRLAM is identical to that of HIRLAM. Thus, the dry-air density for dynamics is calculated using a traditional SL approximation to (4), i.e. not the LMCSL. Therefore, the Enviro-HIRLAM is not formally wind-mass consistent regarding tracer transport. **However, the large scale precipitation fields in the traditional HIRLAM and Enviro-HIRLAM are very similar (see, e.g., Figure 4 in Sørensen et al. (2013)), which suggests that wind-mass inconsistency is of minor importance.**

2. Model applications and validation
a. Sensitivity studies on the model response to aerosol effects do indicate some strong "signals" (difference between coupled and uncoupled runs), e.g. l. 418. But these do not necessarily imply an improved model performance, and the authors should state this clearly in the manuscript, cf. l. 420 "[Korsholm (2009)] found a marginally improved agreement[: : :]", and l. 464-467 "However: : :it is too early to make conclusions about the improvement of precipitation forecasting by implementation of the indirect aerosol effects, because of large uncertainties in parameterisation : : : and due to adjustments of such effects: : :and constants".

Response:
Yes, we agree with the reviewer but don't see contradictions between these statements. Sensitivity studies on the model response to aerosol effects indicate strong "signals", but it doesn't guaranty improvements. E.g., Korsholm (2009) considered evaluations only for some elements (e.g., the coupling interval) in the previous analysis and made corresponding conclusions about the improvements. Other feedback mechanisms effects, especially for aerosol-cloud interactions, studied mostly as sensitivity studies or evaluations for short-term case studies.
The model formulations have only been tested on a case basis and although strong signals have been found this does not imply improved meteorological performance of the model. In particular, testing over longer periods including all seasons was not conducted that time. Furthermore, the interactions between aerosols and the cloud ice-phase are not in a state where improvements would be expected. Therefore we wrote that conclusions about the improvement of precipitation cannot be done at this stage and need more analysis.
Recently similar evaluation studies are realised within the CarboNord project for monthly and annual validation studies. However, they are recently started.

b. This referee believes that careful tuning is needed in view of the large number of parameters in the complex feedback modules, especially with regard to cloud effects. It is not obvious how and to what extent this could be achieved only by comparing final simulation results (i.e., without a further quantitative study of the cloud physical mechanisms themselves).

Response:
We fully agree with the reviewer. The STRACO cloud scheme contains fairly simplified cloud microphysics (heavily parameterized). Hence, tuning is essential for the overall performance of the model, when it comes to precipitation and cloud physical properties. Further work to improve aerosol-cloud interaction and precipitation forecast is needed.

c. An evaluation application for the urbanisation modules was performed for the cities of Paris and Bilbao. There are several issues regarding this application that are neither explained in the text nor in the referenced publications:
i. A domain spatial resolution of 2.5 km appears to be insufficient for such an application.

Response:
Sensitivity tests demonstrated that the 2.5 km was the optimal resolution allowing at the same time to obtain satisfactory reproducibility of the large scale processes and to explore the urban effects at local scale without being diminished due to a coarse resolution, taking into account the limitations of the hydrostaticity of the NWP model.

ii. The resolution of the BEP dense sub-grid is not mentioned. Is it also 2.5 km?

Response: Yes, the BEP is also computed at 2.5 km resolution.

iii. The authors seem having assumed only four urban classes, cf. Figure 10. Such a classification would ignore the important role of green urban areas in UHI evolution.

Response:
Although we assumed four types of urban areas, the urban grid is not fully covered by urbanisation; it also contains a fraction of the green area, defined in the CORINE 2000. The classification and the percentage of urban/non-urban grid can be found in Gonzalez-Aparicio et al. (2010), page 17 Table 4.

iv. Is the 2.52.5 arc-minute resolution (5km) of the AHF data adequate for assessing UHI effects in an urban scale? In the Bilbao case it appears that the entire urban area is covered (and classified) in only 16 cells!
v. Are AHF data constant during the day, or do the authors assume an intrinsic diurnal profile?
vi. Values of 40 or 60 Wm-2 for the AHF are mentioned. Is this a mean annual value or a daily estimation following a seasonal profile?

Response (for iv-vi):
For the Bilbao study, Enviro-HIRLAM didn't implement any AHF parameterization and therefore, the AHF factors were estimated from the LUCY model, as a value for summer and winter without including any daily profile. The value was constant for the urban fraction in the 16 grid cells (it is multiplied – e.g. depends on urban fraction: if 100% -> then max value) covering the area of Bilbao (92 km^2). Although it is not as big as the Paris metropolitan area, the effects of the AHF and the UHI on the atmospheric boundary layer could be visible. A sensitivity analysis of the effects of the AHF and the UHI on the atmospheric boundary layer can be found in Gonzalez-Aparicio et al. (2014).

vii. Concerning the validation process, it is unclear whether a combination of statistical indicators is used or just the correlation coefficient. Not much evidence is presented (e.g. in form of figures or tables) that the model reproduces satisfactorily the mesoscale features.

Response:
The full validation process can be found in Gonzalez-Aparicio et al. (2013) and Gonzalez-Aparicio et al. (2014). The text summarised the overall performance over the two episodes analysed.

viii.It is well documented in the literature that the Paris UHI is expanding just after midnight, but not that this expansion lasts until 11 UTC, especially during a summer period. Comments by the authors would be welcome.

Response:
We agree that the Paris UHI is generally expanding just after the midnight and this is very well documented. In this paper we present the evolution of the UHI on the single day of the 28th July 2009. The UHI was expanding after midnight and the effect was visible up to 11 UTC, not meaning that the expansion lasted until 11 UTC.

ix. Confusion is caused by the fact that in the second paragraph of section 3.2 the authors claim that the model was applied for July 2009, while in the last paragraph of the same section they write ": : :showed that under calm conditions during summer and winter: : :".

Response:
The analysis described in the second and third paragraph of the section refers to July 2009, as the text describes. The last paragraphs indicate the outcomes presented in Gonzalez-Aparicio et al. (2013) and Gonzalez-Aparicio et al. (2014) and focused on winter and summer episodes.

d. Enviro-HIRLAM is operationally used for birch pollen forecasting in Denmark. This appears to be one of the more mature applications of the model, with rather advanced emission, deposition and scavenging modules. However, no mention is made on the effect of online coupling (and the relevant feedbacks) on these simulations. In the conclusions it is mentioned that feedbacks are not important in pollen forecast (l. 711-712). How did the authors reach this conclusion?

Response:
The current version of the birch pollen model presented in the paper has not been used in operational mode yet. Online coupling is important for birch pollen simulations due to dependency of the birch pollen emissions on meteorology. It is also specified in lines 550-552 of the paper.
The current version of Enviro-HIRLAM considers birch pollen as a passive tracer with no pollen feedbacks on meteorology. Online coupling (i.e. impact of meteorology on the emissions) is of main importance in the birch pollen study.

e. Section 3.4 attempts an evaluation of the feedback effects on air pollution forecasting. It is mentioned that online coupling improves the forecast skill, however without referring to specific applications, as for instance the MEGAPOLI Paris campaign.

Response:
As we mentioned, the considered evaluations were done only for some elements (e.g., the coupling interval) in the previous analysis and main conclusions about the improvements were done just for them. Other feedback mechanisms effects, especially for aerosol-cloud interactions, analysed mostly as sensitivity studies or evaluations for short-term episodes. Unfortunately, during the MEGAPOLI Paris measurement campaign we were not able to include measurement studies of aerosol-cloud interactions, so it was not possible to make evaluations of aerosol feedbacks vs the MEGAPOLI Paris data. So, we wrote only about a general reasonable performance of the model vs. measurement data. Corresponding corrections and explanations are made in the revised version of the paper.

From a technical point of view, the paper is excellent. Yet, the authors should check it again for inconsistencies (e.g., both "online" and "on-line" are found in the manuscript).

Response: Thanks a lot. It is corrected in the revised version.