Dear Referee,

thank you very much for your time reading our manuscript and the thorough and helpful corrections and recommendations. Thanks to those, we have improved our manuscripts. Here are some of our additional answers to your requests:

Main concerns:
1. The manuscript often looks like a technical note of the COSMO community where not everything is understandable for a reader not participating to it. So either this manuscript is a technical note of the COSMO or ICON communities and it does not deserve a GMD publication, or it should be revised to avoid this impression.
Thanks to the comments and recommendations, we improved our manuscript tremendously. Hope it looks better now.
1.1 Not all the terms and projects referred within the manuscript are understandable by a reader not used to the COSMO model. Please consider that most of the readers are not part of the COSMO or ICON community and does not know the related projects.
Ex: COPAT project very often mentioned.
We added in the text the explanation of COPAT project (page 7, line 34), we also tried to explain better other terms. Hope it is clearer now.
Ex: R2B8 configurations and other RxBy.
We tried to explain the RxBy better now in the text. A new section „2. General information on ICON-NWP and ICON-LAM“ was added to the manuscript to describe the ICON icosahedron grid and the meaning of RxBy. A new table (Table 1) was added to give some description of the grids mentioned in the paper.
Ex: Figure 1 is incomplete and not completely self-sufficient. Could you please make the text and the figure 1 consistent: ICON-NWP or ICON-ESM are in the text, not in the figure. Same for the Large-Eddy Simulation version for completeness. Please rework the figure 1.
True that we are still inconsistent in the naming of the different ICON configurations. We did not intend to include the Large-Eddy configuration and the ICON-ESM into the family tree. But after re-consideration, we revised Figure 1, it should be complete now. And the texts were also changed to give a more complete description (page 2, lines 23-35).
1.2 Some figures do not have the quality required for a scientific article. It seems that everything was done a bit too quickly without careful final checking by all co-authors.
Ex: figure 2: missing top line. Please defined the PRUDENCE box abbreviations in the Captions
We changed figure 2 and defined the PRUDENCE box abbreviations in the caption. We also revised all other figures. Please also try to zoom in and out a bit, sometimes the borders of the figures do not show off properly on the pdf viewer.
Ex: figure 3: names of the variables in brackets are not standard names. Please use standard names and consistent naming between figures, tables and text.
We changed the names of the variables in Figure 3 to be consistent with other figures and to be more understandable for the readers.
Ex: many figures without units (fig 4, 5, 8, 9, : : :). Please check
The figures were overcropped. Now they should be find and the units are visible.
Ex: some figures with y-axis labels, other without (cf. Figure 7 vs 8). Please check for consistency.
The figures were overcropped. Now the y axis labels are visible.

2. The manuscript focuses a lot on the comparison with CCLM for the evaluation run. I'm not sure that this should be the major point of such an article.
2.1 I understand the will of the authors to focus on this comparison but I would expect that the reader is likely more interested by a fine description of the behaviour of ICLM itself. So at least in the section 4, please spend more time in describing the ICLM behaviour and less time to the description of the CCLM behaviour. You may even want to cut section 4 in two parts, one dedicated to the new model evaluation and the other to the quick comparison with CCLM.
Thanks a lot for this suggestion, we splitted Section 5 into different sub-sections, one is Technical tests and others for Evaluation and comparison with COSMO-CLM.

2.2 In addition, personally I would find more relevant to compare ICLM with all the EURO-CORDEX evaluation simulations performed at 12 km. You can either use the results obtained in Kotlarski et al. (2014) if values are numerically available or recompute some of the key scores using data downloaded from the ESGF. Doing so, you will place the newly-developed RCM within the state-of-the-art of the RCMs in Europe. I know that this request requires massive additional work but I hope that the authors will consider it.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., ... & Nikulin, G. (2014). Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geoscientific Model Development, 7, 1297-1333.

We agreed with the reviewer that comparing ICLM-REF to the EURO-CORDEX simulations is very tempting. However, we do not consider it a must. Taking this publication below for instance:

GIORGETTA, Marco A., et al. ICON-A, the atmosphere component of the ICON Earth System Model: I. Model description. Journal of Advances in Modeling Earth Systems, 2018, 10. Jg., Nr. 7, S. 1613-1637.

They compared their newly introduced model to a precedent model, in their case ECHAM, and not too many others.

As ICON-CLM is not yet well tuned, it would not be really fair to compare ICLM-REF to the CORDEX ensemble. Nevertheless, we agreed to add a comparison using the figures in Kotlarski et al. (2014). A description was added in the Evaluation methods section (page 9, lines 14-22) and a discussion of the results on page 11, lines 3-11.

We plan a more detailed comparision with other models once the model is fully tuned.

2.3 Reading the text, I often feel that the authors are “too proud” of ICLM being so close or even better than CCLM. Again I understand the author point of view after so many years of work and the fear of not being as good as the old model. However the way it is phrased is not scientific (objective) enough and show too much satisfaction with themselves. See for example, conclusion, abstract, page 7 line 28-30. Please re-read the whole manuscript and rephrase keeping in mind that the goal of the paper is to present a first version of the model and not to “kill” the old one. Model developers are often not well placed to judge themselves their new model. At the end, it will be up to the readers and then to the ICLM users to decide if the new model better fit their applications. The future will tell us.

It has never been the authors‘ intention to be too proud or to kill the old model. We re-read the manuscript again and still felt that the wording is quite appropriate, perhaps it is just a habit of language. However, we still did some tone down at some places. Hope that it sounds better now.

3. In a first paper describing a new RCM, I m expecting much more illustrations concerning the technical tests performed with the model before the evaluation run. Many tests are mentioned (time steps, domain decomposition, different computing system) but not really exploited and illustrated. Even if those tests are very appealing to present, for me, they should be at the heart of such paper and each test should be documented by a table or a figure. Currently we need to trust the authors blindly concerning the test results without any proof or trace.

The different time step tests were illustrated in Figure 3 and discussed in the text.

We added a description of the domain decompostion test now to the text (page 9 line 26 to page 10 line 2). But showing figure or table for these tests is difficult since the results are binary identical. In our opinion, the current text clarifies what we tested and how the results are.

An explanation of the different computing system test was also added to the text (page 10, lines 12-16).

3.1 I’m advising to create a section dedicated to the model tests. That is to say to split section 3.1 in two sub-sections, one describing the tests and one for the evaluation run.

3.2 I’m also advising to add at least the “1+1=2” test. That is to say, checking if running 2 months in one job or in two jobs with a restart between the months give the same results or not. This allows to verify the restarting procedure.

Thanks for this interesting advice. We performed two additional tests according to the referee’s recommendation: (1) 2 months in one job without a restart; (2) 2 months in 2 jobs with a restart.

We compared the result of this new tests. The results were identical. A description was added in the text (page 10, lines 8-11).

3.3 Later (not for this specific article), I'm also advising to test the model in the Big-Brother / Little-Brother framework what is for me a mandatory step for any new RCM (see for example Denis et al. 2002)

Denis, B., Laprise, R., Caya, D., & Côté, J. (2002). Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment. Climate Dynamics, 18(8), 627-646.
Thanks again for the interesting idea. We will consider that for our next steps with ICON-CLM.

4. Not enough information on the model configuration and simulation setup. In such article, I'm expecting more information about the model itself and its configuration for the evaluation run. The information given in section 2.1 and in section 3.1 are not complete for me.
4.1 First, clarify what should fit in section 2.1 and what should fit in section 3.1. For me everything general concerning the model itself should go in 2.1 whereas the specific model setup for the simulation (domain, resolution, time step, physical choice, tuning, forcing choice) should go in 3.1. The separation is not always easy but deserve some attention to ease the reading.
We looked at section 2.1 and 3.1 again, in our opinion these two sections are already separated, there is no information about the model setup in section 2.1 and vice versa there is no information about the model itself in section 3.1.

4.2 For the model description (section 2.1), I'm expecting more information and related tables and figures on the horizontal grid (how does the icosahedric grid look like ?), the distribution of the vertical levels, the output procedure (do you output on the icosahedric grid or on a more classical grid ? See the text page 7, line 11). Do you have the option of spectral nudging in addition to the upper boundary nudging? Also add more information about the relaxation zone and lateral nudging procedure (width, variable nudged, strength of the nudging, filtering tricks if any: : :) for example in the paragraph page 4 line 12-16.
We added Table 1 to give some charateristics of the grids that were used and referred to in the paper. A description of the icosahadric grid was added into section 2 (page 3, lines 13-23). We also added Figure 2b in addition to Figure 2a to show the triangular grid R2B8. With regards to the vertical level distribution a paragraph was added in section 2 (page 3 line 24-32). We output on the rotated lat-lon grid not the icosahedric grid, but it can be an option in ICON. This information is added in section 2 (page 4, lines 1-5) and section 4.2 (page 8, line 23).
In ICON limited mode, there are only options for global data nudging or the vertical velocity with the damp layer is damped towards zero. This was already explained in section 2 (page 4, line 13-17).
The information about lateral nudging was added in section 2 (page 4, line 6-12).

4.3 For the simulation setup, I'm expecting there the number of grid meshes for the EURO-CORDEX configuration, the way to define the grid, the numerical cost (compared to CCLM at least), the resolution (explain what R2B8 is) but also the description of the forcings of the run. In particular, in addition to the GHG, SST and sea-ice cover (described in section 2.1), I'm expecting some information concerning the aerosol representation (3D+time variation) that can be very variable from one RCM to another (Gutiérrez et al. 2020), the tropospheric ozone and the evolution of the land-use-land-cover if any (Davin et al. 2020).
Gutiérrez C., Somot S., Nabat P., Mallet M., Corre L., van Meijgaard E., Perpiñán O., Gaertner M.A. (2020) Future evolution of surface solar radiation and photovoltaic potential in Europe: investigating the role of aerosols. Environ. Res. Lett.,15 (3), 034035, https://doi.org/10.1088/1748-9326/ab6666
Davin, E. L., Rechid, D., Breil, M., Cardoso, R. M., Coppola, E., Hoffmann, P., ... & Raffa, M. (2020). Biogeophysical impacts of forestation in Europe: first results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison.

Earth System Dynamics, 11(1), 183-200.
The number of grid meshes for the EURO-CORDEX configuration was added in Table 1. The meaning of R2B8 and general information about the ICON grid was added in section 2, page 3, lines 13-23. The performance in terms of speed with comparison to COSMO-CLM was added in section 5.1, page 10, line 15. We have not optimized the run configuration for ICON-CLM, therefore, we would not further comment on the computing cost at this stage. We expect that after an optimization ICON-CLM computational cost would be much less than at the moment.
Information about the aerosol and ozone climatology we used for our simulations was added in section 4.1, page 6, line 30.

5. A tricky point in RCMs is the capacity to keep or to modify the large-scale information provided by the driving model. Many methods can be applied to check this (Big-Brother/Little-Brother experiment, see above or GCM-RCM temporal or spatiotemporal correlations for large-scale fields often in altitude or cyclone tracking or weather regimes identification). You may want to keep it simple for this study but could you please show at least one illustration allowing to check the lateral forcing procedure? For example, you may want to correlate the Z500 anomaly or the temperature in altitude between the model run and the driver (ERA-Int) at various temporal scales (e.g. yearly, seasonal, monthly, daily, 6-hourly) or anything showing to the reader that ICLM is able to reproduce the large-scale of the driving model at least for some temporal scale (see for example Sanchez-Gomez et al. 2009).
Sanchez-Gomez, E., Somot, S., & Déqué, M. (2009). Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961–2000. Climate Dynamics, 33(5), 723-736.
We calculated the correlation of the geopotential at 500 hPa between ICLM-REF and ERA-Interim data for different time scales (6 hourly, daily, monthly, seasonal, yearly) as suggested. The results are shown in figure 13, a discussion was added in the text as well (section 5.3). In a word, the correlation is pretty high, averaged values higher than 0.925 for all time scales. Correlation is better with longer time scale. ICON-CLM seems to not distort the large scale information of the driving data.

6. Minor comments:
6.1 page 2, line 24: could you explain the difference between “one-way nested subdomain” and “limited-area mode”? For me, it is the same thing. Is it a question of on-line versus off-line?
Yes correct, that is the difference between “one-way nested subdomain” and “limited-area mode”. With “one-way nested subdomain” the nested subdomain and the global domain are being simulated at the same time. Global domain gives forcing to the nested domain, but there is no feedback from nested domain back to global domain.
With “limited-area mode”, there is no global model. The boundaries are simply prescribed from external data. We find that the texts are clear enough, and that the referee could already understand the difference, and plus the “one-way nested subdomain” is not the focal point here, we did not change the text.

6.2 page 3, line 16: for the update of the SST, could we also use lower frequency such as daily or monthly?
Yes physically one can feed SST into the regional model also on monthly basis. But we want a flexible option because technically it is easier to update SST at the given forcing data frequency, so we don‘t have to prepare specifically the monthly data. The text was re-formulated a bit to make things clearer (page 4, line 27).

6.2 page 3, line 19: green house ! greenhouse
We changed from green house to greenhouse in the text.

6.3 page 4, line 6-11: this paragraph could perhaps include more information about the input/output procedure, the file format, the flexibility of the outputs, : : : For example, is it possible to output hourly precipitation and monthly-mean MSLP from the same run? or do you need to output all variables at the same frequency before a post-processing step?
Yes, it is possible to write out different variables with different temporal resolution we added this information in Section 2, page 4, lines 3-5.

6.4 page 4, line 33: grammatical issue
We re-read the line but did not find any grammatical issue. Please re-consider this comment or make it clearer.

6.5 page 4, line 24-30: could you explain more the restart procedure and the job management and its flexibilit? Could you perform daily run, monthly run, yearly runs ? Or do you have a mandatory time slice such as one month?
Yes one can run the model for a wished time period (that is why we could be able to do the 1+1=2 test with two months in one job without a restart). But we normally choose calendar month. Thanks. We added this information to Section 3.2, page 6, lines 6-7.

6.6 page 5, line 2: could you tell more about the tuning strategy for ICLM. What do you try to optimize?
What we meant with these text is to introduce the Starter Package of ICON-CLM as an useful tool for different purposes. One is using the Starter Package in tuning ICON-CLM. It is not our intention to tune or to set the strategy of tuning ICON-CLM.
This work is planned in the next phase of COPAT project and will be introduced later.

6.7 page 5 and in many places: EU-CORDEX ! EURO-CORDEX
We changed in all places to EURO-CORDEX.

6.8 page 5, line 20: 30 km. Give also the value in hPa.
Yes we gave now the value in hPa (page 6, line 26)

6.9 page 5, line 29: give the list of the variables nudged and the nudging coefficient
We added this information on Section 4.1, page 7, lines 5-6.

6.10 page 5, line 31-33: The use of many unexplained grid names (R2B8, R3B8, R3B7) is confusing. Simplified or explain. Also in the paragraph, you mention tuning parameters from global settings but setup from LAM: : : clarify
We added the description of the ICON grid in which the names and denotes of the grids are explained in Section 2. Also Table 1 gives information on the mentioned grids.

6.11 page 6, line 8: could you compare the 120 s time step with state-of-the-art RCM time steps at the same resolution?
We do not have information about the time steps from other RCMs, they are also not stated in Kotlarski et al. (2014). Time step of COSMO-CLM at 12 km is 100 s, this information is added on page 7, line 28.

6.12 page 20-23: is the reference CCLM simulation published? Any reference to refer to? If yes, cite it. If not, you need to describe it in the method section or to use a published run such as one of the EURO-CORDEX evaluation simulations performed with CCLM and available on the ESGF.
The run CCLM-REF is unfortunately not published yet. But it was done with the most recently recommended version and configuration of COSMO-CLM. That's why we chosed this run and not one in EURO-CORDEX evaluation simulations.
The description of the CCLM-REF is in section 4.1, page 7, line 32 to page 8 line 4.

6.13 page 6, line 34: clarify that you are considering only land points.
Yes we clarified this in Section 4.2, page 8, line 27.

6.14 page 7, line 12: typing issue?
Yes indeed. The letter „f" is missing in front of the „or". We corrected now.

6.15 page 7, line 21: In your case, if I understand well, the RMSE measures a skill related to temporal variations of the variables over the PRUDENCE boxes. So I would have dedicated STDEV to a spatial skill score by averaging in time before computing the standard deviation. Currently STDEV is spatio-temporal score if I understand well, what is therefore quite difficult to interpret. Please, consider to change this. Also table caption mentions "spatial standard deviation" whereas the text mention "spatiotemporal

standard deviation". Please clarify.
Yes, the sdtdev was spatio-temporal score. Thanks for the advice, we changed it to spatial by calculating the time average and then the deviation. The text was also changed to „spatial". The description of the STDEV was adapted (Section 4.2, page 9, line 10-13). The discussion of the results was also changed accordingly.

6.16 page 7, line 21: For the quantitative score, I'm not forcing you to do so but it could have been a better option to plot Taylor diagrams (incl. RMSE, correlation, standard deviation) in order to be more exhaustive in the evaluation of the runs: for example a spatial Taylor diagram per season for all European land points and a temporal Taylor diagram for each PRUDENCE box. This is just an advise. In particular, it allows to put all boxes or all seasons or all variables on the same figure.
We did actually make the Taylor plots at the beginning, simply because they are part of the Evaluation tool of ICON-CLM and are made automatically when the tool is run. But then we decided to present our results in the form of the tables. Perhaps it is a matter of choice.

6.17 page 7, line 21: If you decide to keep the score STDEV, I propose to put in the tables the ratio of the standard deviations (Model/Obs) in order to have only 2 columns as for the RMSE allowing to easily see the best model for every line.
We changed the stdev according to the referee's comment.

6.18 page 7, line 28-30: this small paragraph illustrates well my major comment 2.3 with terms such as "very good performance", "consistent for all six evaluated variables", "already of similar". Please rephrase in a more objective and scientific way without overstating the results obtained. Also remember that the ICON project started 20 years ago. So the model is not so new and has been already tuned and adapted at that resolution over the European domain. I'm aware that a model used in climate mode
can show biases not seen in weather forecasting mode but still, you are building on the weather forecast experience. Also note that the model performance is not "consistent" for all variables. From my point of view, it seems better for temperature-related variables than for precipitation or MSLP. Here again, a section comparing the ICON run with all the Euro-CORDEX RCM runs in evaluation mode would be more conclusive (see previous major comment).
The mentioned small paragraph was removed and replaced.

6.19 page 8-9-10: Please reorganise the text of those sections to put first the description and discussion of the ICON biases before comparing more quickly with the CCLM reference as the reader want more information about the strengths and weaknesses of ICLM and less about CCLM. Currently I find that the ICLM description is too light and the CCLM description too fat.
The manuscript was changed heavily. We hope the ICLM part is better now.

6.20 page 8, line 10: "no bias" ! When the median bias is near zero, it does not necessarily mean "no bias", it can mean "bias compensation in space". Rephrase.
In the text, we wrote „nearly no bias" as the medians were about 0.01 K and the percentile boxes were pretty short.
We re-formulated to „relatively small bias" now (page 10, line 28).

6.21 page 8, line 17: "extreme daily temperature" ! avoid to use the world extreme for min and max daily temperature. It is misleading for the reader as "extreme" is often kept for specific statistics or indices. Check everywhere. Also page 9, line 2, line 3.
Thanks for the comment. We checked everywhere in the manuscript and replaced the word „extreme" with other expression.

6.22 page 8, line 21: "the bias was larger". All the text of the results is written at the past form. I'm not an English specialist but it would be easier to read at the present form ! "the bias is larger ...". Please consider to change this everywhere in the results section.

The authors are also not English specialists. But from the teaching of tense in academic writting, for results section, past tense is used to describe results obtained. Simple present is used to describe figures, tables. So we would like to keep the tense as it is for the moment.
The manuscript will go through English editting once it is accepted.

6.23 page 9: please state that ICON is not so good for Summer day statistics. I don't understand why the representation of figure 11 is not similar to the representation of figures 5 to 10 with a box plot representation. A black box can be used for the observation in addition to the green and blue boxes in that case.
Yes we added in the text that ICLM-REF is not as good for summer day as for other temperature-related indices. But still it is beter than CCLM-REF in 3/8 sub-regions (ME, SC, AL) and on average over the whole Europe (EU). In three sub-regions, both experiments were equal (BI, IP, FR).
In all figures from 5 to 11, blue is ICLM-REF, green is CCLM-REF, and in figure 11 black is observation. In figure 5 to 10, there is no black because they show already the biases. We could add black boxes representing observation in figures 5 to 11, but since each sub-figure has already 16 boxes, we do not want to make the figures too crowded. We would like to keep figure 5 to 10 as they are, with the biases.

6.24 page 9, line 19: not sure I agree that CCLM overestimates the precipitation. It is relatively well balanced over Europe contrary to ICON.
As stated throughout the manuscript, CCLM-REF was better for precipitation. But still it simulated more precipitation in comparison with E-OBS (of course E-OBS properly has measuring error, see the next comment). This overestimation can be seen in figure 4, and even more obvious in figure 11.

6.25 page 9, line 20. Please cite a reference for the "too low values". For precipitation, please also mention and discuss the strong model biases over the topography.
We added a citation for the low values due to gauge undercatch and evaporation. The text was also revised a bit too make it clearer what we meant (Section 5.2.2, page 12, lines 26-28).

6.26 page 9, line 29: "summer had the smallest variations". Not so true if you think that precipitation is very low in summer for some regions. Computing the error in % (even without showing them) may help for discussing the results
That is true that due to the low precipitation values in summer, the variation of the values are lower than in other seasons. But the statement about the bias was correct.

6.27 page 9, line 34: "five out of height" ! for me it is 7 out of 8. Please check in table 6.
The stdev was replaced by the stdev ratio (model/obs). The text was changed accordingly, it is six out of eight.

6.28 page 10, line 3-6: for me by eye, CCLM-REF seems better than ICLM-REF for those indices. Please re-assess.
For Wet days index ICLM-REF was better in 6/9. Heavy precipitation 4/9, very heavy precipitation 3/9. That makes 13/27, and few times the 2 models are give almost the same results. We would keep the same statement that none of the models is better than the other for these indices.

6.29 page 10, line 7: please cut the MSLP and cloud section in two sections, one for each variable for consistency and add the Table 8 for the cloud cover again for consistency.
We cut the MSLP and cloud into two sections (Section 5.2.3, 5.2.4) and table 9 was added for cloud cover.

6.30 page 10, line 8-11: any explanation for the MSLP biases in both models?
We do not know. ICLM-REF has a bit higher MSLP than CCLM-REF which shows in its positive MSLP biases, compared to the negative biases from CCLM-REF. The pressure pattern from the driving model is quite well kept in both simulations though.
An additional evaluation for MSLP was added (Figure 12). In this figure, we compared MSLP from ICLM-REF and CCLM-REF to that of the driving model ERA-Interim.

6.31 page 10, line 9: for MSLP, it seems that the biases can reach values higher than 2.5 hPa (cf. Figure 4 over Spain for ICLM.

It is true. The text was changed in page 13, line 22.

6.32 page 10: same question for the cloud biases. Any explanation or hypothesis?
We just know from ICON that it produces a bit too little cloud and therefore has a positive bias in radiation. Perhaps this underestimation of cloud is stronger in COSMO-CLM. But we do not know for sure about COSMO-CLM, therefore would not add further comment on this.

6.33 page 10, line 20: not clear where you find the +/- 5% values
This line discusses bias of cloud cover in Figure 4. On ICLM-REF side the colors are mostly light pink/blue which correspond to -0.05 to 0.05 and translate to +/- 5%. We added the explanation in the text to make it clearer (Section 5.2.4, page 14, line 11).

6.34 page 10, line 20-21: "overestimation of the cloud cover : : : cold bias". ok for the causality for tasmax but this is often the opposite for tasmin. Rephrase.
What is refered to here is tas. Looking at Figure 4, the overestimation of cloud cover aligns with where CCLM-REF has cold bias for 2 m temperature. We added the reference to Figure 4 to the text to avoid confusion (Section 5.2.4, page 14, line 14).

6.35 page 11, line 11: Personally my assessment is that CCLM is better than ICLM for precipitation. Please re-assess. I agree that models are equivalent for MSLP.
It is hard to say from the areal average since some part ICLM-REF is better and some part CCLM-REF is better. But we agree that overall CCLM-REF is better with precipitation and editted the text (Section 6, page 15, line 23)

6.36 figure 3: please make this figure easier to read. For example by increasing the thickness of the curves? Possibly showing only seasons or showing maps? Try to make it simpler and more informative with the key message easier to catch for the reader.
Our idea with this Figure 3 is to test whether ICON-CLM gives remarkably different results due to the choice of time step, which happened with COSMO-CLM. The figure shows that the lines are quite close to the others and no line stands out of the bunch. We would like to keep this figure like that. Of course as written above, we changed the name of the variables and made the sub-figures a bit nicer now.

6.37 figure 4: showing the areas where the differences are statistically significant or not may lead to a more informative figure and make it easier to describe in the text in order to focus only on signaificant biases. Please revise the map projection for figure 4 (it is ugly currently) in order to limit the zone without information in each panel. Also "shave your model" that is to say remove the relaxation zone or comment the model behaviour there in the text.
We revised the map projection and removed the relaxation zone.

6.38 figure 5-10: I like such figures. Check the y-axis labels and the units everywhere.
Yes the figures were overcropped. Now y axis labels and units are visible.

6.39 Table 1: Please add more information about the physics by splitting the deep convection and shallow convection lines and by splitting the radiation in short-wave and long-wave radiation. Add in this table all useful information and references for the physics as it will likely serve as reference for many articles afterwards.
We added more information regarding shallow/deep convection and short/long wave radiation as requested.

6.40 Table 8: please add a table for the cloud cover

We did. It is table 9.