egusphere-2025-4726
Review1 of "Optimisation of ICON-CLM for the EURO-CORDEX domain: developments, sensitivities, tuning"

Thanks to the first reviewer for the extensive work on our long paper! We appreciate your time and willingness to follow all our steps and conclusions; we also appreciate very much pointing out missing details and suggestions for improvements that help the reader in their understanding. We addressed all suggestions, answered your questions, and updated the manuscript accordingly. We split your original review into small paragraphs (in black) for direct answers (in blue):

General evaluation:

This study systematically investigated the parameter sensitivity in the ICON model and applied a structured optimization strategy to determine the optimal configuration for ICON-CLM over the EURO-CORDEX domain. The authors present the results from both the expert tuning and the Linear Meta-Model optimisation (LiMMo) method, and the LiMMo method effectively reduces model biases. Specifically, when compared to the default ICON NWP setup, the optimized configuration significantly mitigates underestimated surface air temperature, overestimated incoming shortwave radiation and overestimated latent heat flux over the sea. Based on these improvements, the authors propose this setup as the standard ICON-CLM model configuration for the EURO-CORDEX domain.

The paper provides a clear description of the tuning strategy and its primary findings. I recommend this paper for publication in GMD after the authors address the following comments, which are primarily aimed at providing further clarification.

Major comments:

1. Enhancing clarity (ScoPi and AMSESS)
To ensure the manuscript is a stand-alone piece, additional technical detail regarding the ScoPi and AMSESS metrics is needed, particularly as the primary reference (Geyer, submitted) is currently inaccessible to readers.
The preprint version of the draft by Geyer (2026) is not yet available, although we submitted the revised version on the 14th of March. Therefore, we attach the submitted pdf ensure access for the reviewers.

- Please include the core equations for ScoPi (e.g., the equation cited in L429) and AMSESS (L406). Providing these within the Methods section will significantly improve the paper's readability and allow readers to evaluate the results without external searching.
The equation (Eq. 8) for AMSESS is added in Sect. 2.6:

$$\mathrm{AMSESS} = \begin{cases} 1 - \sigma_{ts}^2 \cdot \sigma_{ref}^{-2} & \text{if } \sigma_{ts}^2 \leq \sigma_{ref}^2 \\ \sigma_{ref}^2 \cdot \sigma_{ts}^{-2} - 1 & \text{if } \sigma_{ts}^2 > \sigma_{ref}^2 \end{cases},$$

The 3 equations (Eq. 9-11) for ScoPi (from Geyer, 2026) are added in Sect. 2.6.1:

$$\mathrm{ScoPi} = \begin{cases} F_{ss} + 0.5 \cdot \mathrm{sign}(F_{ss}) & \text{if } |F_{ss}| > 0.5 \\ F_{ms} & \text{otherwise} \end{cases}$$

$$\mathrm{ScoPi}_{region}(r) = \sum_{v,m,s} c_v \, \mathrm{ScoPi}_{v,m,s}(r) \, \mathbf{1}(|\mathrm{ScoPi}_{v,m,s}(r)| > 0.4),$$

$$\mathrm{ScoPi}_{simulation} = \sum_{r} c_r \cdot \mathrm{ScoPi}_{region}(r),$$

- In Table 2 and L440, the text mentions that “different weights are considered per variable.” Please explicitly define how these ScoPi-weights are determined for each variable to ensure the methodology is reproducible.

With the given equation for ScoPi_{region} it is defined how to apply the ScoPi-weights. The weights are user-defined. However, we tested the dependency of our method on the setting of weights and found reproducible results for varying weights (Geyer, 2026).

- To assist the reader in interpreting the results, please clarify the physical or statistical significance of the ScoPi magnitude. Does a higher value represent superior model performance? I suggest briefly introducing this interpretation in the Method section…

We added at the end of ScoPi explanation: “Tests with several metrics done by Geyer (2026) revealed that the ranking between the tested simulations remains stable in relation to the reference: the higher the ScoPi score the better the overall performance of the tested simulation.”

…and include a reminder in the relevant figure captions (e.g., “Higher ScoPi values indicate...”) to guide the reader through the results.

The figure caption of Fig. 15 is prolonged by the sentence “Higher ScoPi values mean that the test simulation is more consistent with the selected observations than the reference simulation.”

2. Implementation of allow_overcast (L231)

The technical implementation of the `allow_overcast` parameter is currently unclear. Specifically:

- On what timescale is this parameter modified (e.g., monthly)?

We did different tests (see Table 6):

- ALLOW_OVERCAST-PAR: constant allow_overcast
- ALLOW_OVERCAST-PAR(m): allow_overcast has monthly predefined values.

We changed “passed the appropriate value to the namelist.” to “passed the appropriate monthly value to the namelist.”

- How are the user-defined deviations from the mean `aoac` estimated?
We analysed monthly biases of simulations with first constant values of *ac* to adjust the impact accordingly per each single month leading to *t1* and in the further work to *t2* to *t4* (as listed in Table A2).

- Is `aoa` also a tunable parameter?
Yes, it is listed later on in Table 9 with the margins 0 to 1.5 where the value 0 would lead to the constant value of ao for allow_overcast.
We changed the text from “, scaled with an amplitude aoa” to “, scaled with a tunable amplitude aoa”

- Please elaborate on the physical or modeling rationale for introducing a monthly dependency for this specific parameter.
The development of *t1* to *t4* was based on in-depth analysis of monthly model biases. With tests, we found out that the yearly cycle *t4* revealed the best results.
We added: “This change was inspired by the identification of monthly variations in model errors through a detailed analysis of test simulations.” to the text.

3. Criteria for Parameter Sensitivity (L500)
There appears to be a discrepancy between the text and Figure 2 regarding "non-sensitive" parameters. For instance, `ecrad_llw_cloud_scatt` (1.8 on psl in JJA) and `czbot_w_so` (1.9/2.0 on tas/tasmax in JJA) show metrics significantly higher than 1.0. Please define the specific threshold or criteria used to categorize parameters as "sensitive" vs. "non-sensitive" to ensure consistency with the visualized data.
In L501 we stated: The remaining parameters tested exhibit a sensitivity around twice the intrinsic variability or larger, especially during the summer, and are classified as "sensitive parameters".
We added in the methods section 2.6.2 the statement: “We consider the model to be “sensitive” to a parameter change for $SENS^{seas}$ values above 2.”

Minor comments:

1. Accessibility of Model Configurations
While Table C1 provides a comprehensive summary, the frequency with which configurations like C2I200c, C2I250c, C2I268c, C2I291c and C2I294c are discussed makes it difficult for the reader to track their differences throughout the paper. I would suggest including a condensed table in the main text summarizing these key configurations to provide a quick reference for the reader during the discussion.
At the beginning of Section 3.5 the key configurations are explained (C2I200c, C2I250c, C2I268c, C2I291c and C2I294c). We added the brief description of these configurations as a row in Table 10 to improve the overview.

2. Parameter descriptions (L193-201)
The tuning parameters require slightly more context in the main body to be fully understood:
- Please clarify that `rat_lam` applies specifically to latent heat flux over land.
“rat_lam influences the latent heat flux over land only.” is added to subsection 2.2.2.

- Provide a brief definition or functional description for `cr_bsmin`.

"cr_bsmin is the tunable minimal bare soil resistance for evaporation." is added to subsection 2.2.2.

- Ensure an explicit cross-reference to Appendix A is included in this section.

Detailed cross-referencing was suppressed by the publisher (as done by us i.e. for aoac* in Tab. 6 to Tab.s A2 entry for aoac*, so that no parameter-based links to the entries in Table A1 are possible here.
We added "All namelist parameters used are listed in Table A1." at the end of the paragraph.

3. Physical drivers of radiation changes (Figure 7 & L559)
The paper notes an increase in downwelling longwave radiation in Figure 7. Please add a brief discussion on the underlying cause. For example, is this change mainly driven by shifts in cloud cover, or are other factors involved?
Cloud cover differences between the two simulations indeed have a diurnal cycle over northern Africa (see Fig. R1 below). A maximum cloud cover difference of about 1 % can be observed at about 20 UTC. Investigating temperature and relative humidity differences at different levels, we found that the temperature in C2I105 (with the higher AOD from MACv2-SP) is by about 0.5 K higher in the lower troposphere (at levels of 925 hPa and 850 hPa, see Fig. R2 upper row) and lower at 500 hPa, which explains the relative humidity difference at 500 hPa. The cloud cover difference must therefore be maximum at about 500 hPa (Fig. R2 bottom row). The emission of thermal radiation from these mid-tropospheric layers is relatively small. Thus, the cloud cover differences can only partly explain increased nighttime warming.
On the other hand, air layers with a high aerosol load also absorb heat and can influence heating rates.
Additionally, the switch from Tegen to Kinne aerosol in ICON also comes with a different treatment of aerosols. While Tegen provides AOD at 550 nm for different species, the Kinne climatology directly provides optical properties, which can result in a very different optical properties in the end In the article where Stefan Kinne presents sensitivities of his climatology, he describes a positive aerosol radiative forcing over Africa, which confirms that the Kinne and thus, the MACv2-SP climatology lead to a warming in that region.
In the text, we added:
"Increased rlds can partly be explained by increased cloud cover at night. However, the cloud cover difference is small and cannot be detected below 500 hPa, so the effect on thermal radiation is minor. Kinne (2019a) clearly shows a warming due to direct radiative effects of the MACv2 aerosol over northern Africa and Arabia. The explanation is that the "mineral dust aerosol particles in those regions are relatively large, elevated (off the ground) and absorbing." With that, mineral dust can "contribute to a significant greenhouse effect". However, it is debatable if MACv2 overestimates the increase of longwave radiation as it neglects the natural variability of mineral dust both in terms of variability of size distribution and mineralogical composition, which can have a considerable influence (e.g. Sicard et al., 2014; Gao et al., 2026, and the references therein)."

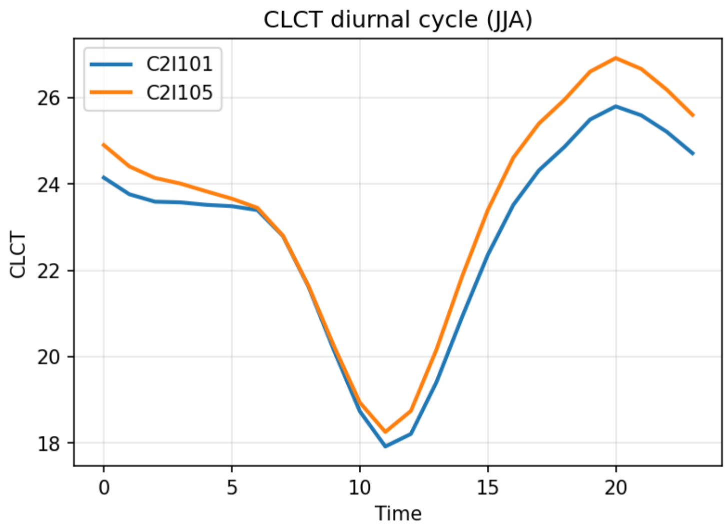

Figure R1: Summer mean (JJA) hourly cloud cover CLCT [%] of the subdomain indicated in Figure R2 for the simulations C2I101 (Tegen aerosols) and C2I105 (Kinne-SP aerosols).

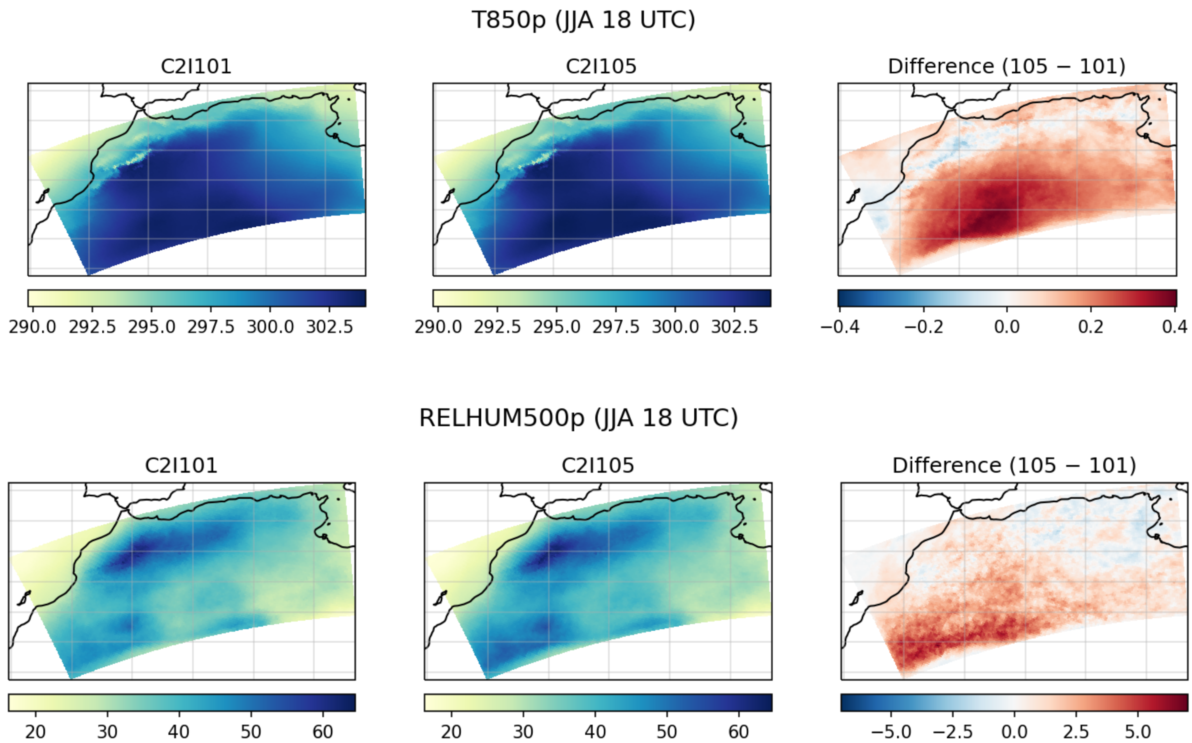

Figure R2: Summer air temperature [K] at 850 hPa (upper row) and relative humidity [%] at 500 hPa for a selected subdomain for the simulations C2I101 (Tegen aerosols) and C2I105 (Kinne-SP aerosols) and the differences between them.

4. Visualization of parameter impact (Figure 15)

To improve clarity, I suggest labeling the y-axis with the “parameter names” rather than the “Simulation IDs”. Since the discussion centers on the physical impact of specific parameters,

this change would make the figure much more intuitive.
Done.

5. Consistency in comparison (Figure 21 and 22): In Figure 21, the authors use `C2I250C` for comparison, but switch to `C2I268C` in Figure 22. Is there a specific reason for switching the reference configuration between these two figures? Please clarify this choice in the text to ensure the comparative.
The primary motivation behind selecting specific configurations in Figure 21 and Figure 22 was to show the 2D maps of variables with the strongest bias change, given in Figure 20. This is stated in the beginning of Section 3.5.3: “In the current section, we show the seasonally averaged 2-dimensional bias plots for the most significant changes identified in the RMSE analysis (see Fig.20 in Sect. 3.5.2)” where we added “Fig.20 in” to help the reader to follow.

Technical corrections:

L4: “RCM” appears for the first time without a definition. Please provide the full name here.
We added: “regional climate model”

L14: Suggest changing “the new configuration” to “the optimization configuration” and “revised external datasets” to “updated higher quality external datasets” for clarity.
‘optimised’ added.
We replaced “revised” with “updated higher quality”.

L60 & L68: World Climate Research Programme (WCRP) is defined at L60. Please use the acronym “WCRP” directly at L68.
Done.

L90 & L213: Correct “parametrisations” and “parametrised” to “parameterisations” and “parameterised.” Please ensure this spelling is consistent throughout the entire manuscript.
We have chosen British English, which makes colour, ~ise and kilometre, etc., correct - we searched for misspelt words and corrected them. Parametrisation and parameterisation are both valid to our knowledge, with parametrisation within a mathematical/physical context.

L105: Please provide the full name for RMS (Root Mean Square?) at its first appearance.
“Root Mean Square” is added to the text.

L110: Change “grows” to “grow.”
The sentence “As the insight on … grows during the optimisation process, the tuning aim may be iteratively revised.” is correct with grows.

L136: Correct “hon-hydrostatic” to “non-hydrostatic.”
Done.
L145 & L174: Change “icon” to “ICON.”
Done.

L308: Please clarify the intended meaning of “climatology of ECMWF.”
The sentence is changed to “.. adjusted with a satellite-based CDNC retrieval by Bennartz and Rausch (2017) and Grosvenor et al. (2018) as used in ECMWF’s IFS model.”

Eq.3: Please define “m.”
“are the test parameter and reference parameter values” is extended to “are the test parameter and reference parameter values for the parameter with index m “

Eq.8: Please define “\delta_pm”.
Before Eq. 8 the following explanation is added: ”Each sensitivity experiment simulates the change of only one parameter from the reference value p_m^{ref} by an increment of Δp_m.”

L369: Could the authors clarify the rationale for interpolating observational data to the model grid? If the observations are coarser than the model resolution, please justify this approach.
Generally, we share the opinion of the reviewer that differences should be calculated on the coarser grid. E-OBS and ICON-CLM have about the same resolution of approximately 12 km, so the method is not sensitive to the interpolation. However, HOAPS has a resolution that is approximately four times coarser than the model, so it would be better to conduct the error norm computation on the HOAPS grid. Nevertheless, the reduction in the bias of the latent heat flux over the sea was confirmed by the independent analysis presented in Section 3.5.3, where the comparison was performed on the observational grid. We will address this issue in the manuscript (around lines 369–370) by explaining that the RMSE computation should ideally be performed on the coarse grid after the area-weighted averaging of fine-grid data.

Table 5: Suggest rephrasing “The signal column defines the parameter signal” into “... denotes the two simulation IDs (without 'C2I' prefix) that are used to estimate the impact of parameter changes.”
Done.

Figure 4: Ensure “HWSD v2.0” matches the naming convention used in the Figure 3 caption.
The abbreviation is changed to the officially given “HWSD v2.0” in the text and the figure.

Additionally, please explain the notation “C2I232 - C2I230” in the caption for clarity; this suggestion applies to all similar figures.
In-Figure captions as “C2I232 - C2I230” denote the ID’s of the used simulation to calculate the signal.

L543: Based on Figure 6, “do not significantly affect” appears to be a more accurate statement regarding the changes in subgrid slopes.
We chose the wording because we analysed with inspection of Fig. 5c where the slopes differ in the two data sets. The shape of these regions (e.g. Pyrenees and Alps) is not visible in the sensitivity plot, Fig. 6 a).

L585: missing comma between 'itype_z0=2' and 'z0.'
Done.

Figure 9: There appears to be a discrepancy in the spatial coverage; the upper panel shows signals over the sea, but the lower panel includes signals over land. Please verify if this is intentional.

Figure 9 shows the latent heat flux at the surface (hfls) for winter in the upper row. The left panel shows the signal for rlam_heat (reduced from 10 to 6.25), which is calculated as sim. C2I205 − reference sim. (C2I200) for the whole domain; only some land points, e.g. over Africa, show value outside the intrinsic variability range of ±2 W/m^2. The same holds for rat_sea on the right side. The signals for the summer months cover ocean and land as well.

Figure 9 (caption): The term "minimum resistance" is used here but does not appear in the Appendix table. Please ensure the terminology in captions matches the parameter tables and main text exactly.

The entry in Table A1 for rlam_heat is "The larger rlam_heat, the larger is the laminar resistance." So, higher rlam_heat values mean higher resistance and therefore reduced turbulent heat flux.

The figure caption is changed to "Figure 9. Mean differences (1980-1984) of hfls for reduced ~~minimum~~ resistance to turbulent heat fluxes rlam_heat."

Figure 10: The colorbar title "rsdu" should likely be updated to "rsus."

Done.

L676: Please clarify if "inwp_cldcov" is the same as "inwp_cldcover" and use consistent terminology.

inwp_cldcover is the right name. Typos are corrected.

L678: Missing space between "2.6)" and "on".

Done.

L683: Clarify "bottom height" as "cloud bottom height."

Done.

L715: Please check if "reference: -0.05" should be "0.05."

Corrected to 0.05.

L717: Please clarify the physical meaning of the phrase "for rh close to 100%."

We have rewritten the paragraph describing the effect of the parameter. We introduced a reference to the equation showing the parameterisation RH(tune_box_liq) and describe the tun_box_liq parameter more clearly.

Figure 15 & 19: There is mixed use of ScoPi and ScorePEvi in the captions and axis labels. Please standardize these.

x-axis labels are changed to ScoPi.

L752: Change "since not" to "since it does not".

'it' is added.

L762: Change "on in the second row" to "in the second row".

'later on' is deleted.

Figure 18: The label “swd” in the third column should be changed into “rsds” for consistency with the rest of the text.
Done.

L853: “Under consideration for consistency” - please clarify what these are meant to be consistent with.
The two parameters, tune_albedo_wso(1) and tune_albedo_wso(2) belong to the albedo correction according to Eq. 1 and its explanation. Therefore, it seems to be consistent to include both parameters in the optimisation. The resolution of the optimisation led to values near the tested values of -0.1, which confirmed our assumption.

L938: Change “is resulting in” to the more direct “results in.”
Done.

L942: Missing space between “tas,” and “tasmin.”
Done.

L999-1000: This sentence (“The ICON-CLM simulations with ...”) is currently unclear; please rephrase it to better convey the intended meaning.
Now we change “The ICON-CLM simulations with LiMMo-derived configurations confirmed the bias reductions found by the meta-model.” to “The analysis of the ICON-CLM simulations with parameter settings distinguished by LiMMo confirmed the bias reductions we expected through the ‘forecast’ by the meta-model.”

Review2 of Geyer et al. “Optimisation of ICON-CLM for the EURO-CORDEX domain: developments, sensitivities, tuning”

Thanks to the second reviewer for the work with our long paper as well! As for the first review we appreciate very much your time and willingness to follow all our steps and conclusions. We analysed your suggestions and answered all your questions and updated the manuscript accordingly. We split your original review in small paragraphs (in black) for direct answers (in blue). Sometimes we refer to the answers given for the first review, ‘double requests’ motivated us even stronger to improve our manuscript:

Summary and Overall Comments:

The authors describe their approach for improving ICON-CLM (for a regional domain), with a particular focus on integrating more objective tuning efforts into the model optimization approach (which is exciting to see). Up to now, a more automated tuning method has not been used, and this addition is valuable and in line with broader community efforts to make tuning more objective. With respect to the model, the authors are particularly aiming to reduce overestimation of incoming SW and too low 2-m temperature biases as part of overall optimization efforts. The paper is mostly OK as written (**recommend somewhere between minor and major revision**), and I hesitate to recommend more as it is already long (with ~20 figures and ~10 tables), but I do think it is worthwhile to **clarify throughout exactly how they view or partition**

tuning into “expert” and “objective” components because even objective tuning requires decisions on parameters and metrics provided by experts (the ‘computer’ itself does not decide on parameters and metrics, and thus the results are human dependent too). I think of tuning from the perspective of advances in automation and AI enabling more parameters and metrics to be considered (with the whole process more efficient), but with experts always involved in deciding on which parameters to tune & metrics to tune against -- I do not think tuning is fully separated which the paper seems to maybe imply as a new approach (although perhaps that is accidental, and only minor wording changes are needed throughout?). Additional general comments/questions are laid out below.

General Comments:

1. Figure 1 is very helpful, but it contains quite a bit of text and acronyms, and I am still not certain I understand the formal steps at arriving at tuning that is a balance of expert judgment and objective tuning. Existing automated tuning (or auto-calibration) efforts are not trying to silo the two efforts, necessarily. **Are the authors proposing a new way for merging the two aspects of tuning that CURRENT autotuning or autocalibration efforts are not considering?** At present, as it exists, autotuning still implies expert involvement (because someone has to choose the parameters and metrics they compare against) – thus, the comment around line 892, and elsewhere – “LiMMo tuning cannot achieve the same model quality as expert tuning” – sort of assumes that LiMMo is determining things on its own, but if you coded up the same features that the experts are looking at, couldn’t LiMMo do better? **I do not fully understand the implied extent of separation between experts and objective tuning as being referenced throughout.**

We thank the reviewer for the comment and related question. We agree that it has not been made clear what method was applied and what the conclusion from this study is with respect to the method suggested. Accordingly, we revised the description of the tuning concept, the analysis of the simulation results of optimised configurations and the conclusions.

We revised the last paragraph of Section 1 (Introduction), the first paragraph of Section 2 (Methods) and the first and last paragraph of Section 2.1 (Tuning strategy) to highlight that expert and LiMMo tuning are different tuning methods, that their results are independent and thus alternative methods of configuration optimisation and that LiMMo tuning can be used as an alternative option to expert tuning. Furthermore, we made it clearer that stage 3 in Figure 1 has been part of the workflow of this study and that following this procedure, the results of expert and LiMMo tuning can be compared with each other.

We revised Section 3.5.1 to highlight the differences between the ScoPi analysis and the norm optimised by LiMMo to avoid misinterpretation of the tuning results.

Finally, we revised paragraph 1 of Section 4 (Conclusions) to make it clear that LiMMo can seamlessly replace or complement expert tuning.

2. In the tuning strategy – **it is mentioned that the tuning effects of combined parameter changes involves experts – how exactly is this done?**

We agree that it has not been made clear how the expert tuning is applied.
We revised in Section 3.3 paragraphs 1 and 2 and described in more detail the expert decision process resulting in new reference configurations. In Section 3.4.1, we revised paragraph 1 and described the expert tuning procedure in more detail. In Section 4 (Conclusions), the paragraph "Third, …" has been revised, and the conclusions for expert and LiMMo tuning are described more clearly.

3. **How is structural error in the model accounted for?** How is it ensured that parameter tuning is not compensating for structural errors or is a possible compensation not a concern in their modeling efforts/goals?

We thank the reviewer for this fundamental comment, and we would really like to answer the question. However, it is beyond the scope of the paper to answer it. The tuning concept presented cannot distinguish between structural error, also named unphysical model behaviour and physically justified parameterisations of e.g. subgrid scale processes.

The author's opinion is that all models used for operational NWP, all RCMs and ESMs, aim at reproducing the behaviour of the real system quantitatively. All these models simulate the correct behaviour (where applicable) through parameterisation. For example, they underestimate the surface height variability. Nevertheless, they are designed to simulate the observed orography-driven precipitation.

The authors do not know how to distinguish between physically well-justified parametrisations of subgrid scale processes (e.g. generating additional precipitation) and physically inconsistent parameterisations (structural errors) generating additional precipitation.

In this sense, it may be that the optimised configuration is compensating for structural model errors. However, this cannot be distinguished in general, at least the authors do not know how.

4. For the observations used in the tuning**, is there any concern about observational uncertainty and its role in the parameter tuning process?** Are all the observations, many of which are retrievals, "truthful" enough or would accounting for uncertainty in the process influence the parameter settings?
Thank you very much for giving us food for thought, but it is not in the scope of the paper to analyse the quality of the observational datasets. The E-OBS data set relies on observations which are contributed from many institutions. Shortcomings are known and listed on the website (E-OBS gridded dataset). Where these shortcomings might disturb our method, e.g. for the surface pressure, we used reanalysis data additionally as the next reliable source for Europe-wide data. Where no station-based data exist, meaning over the ocean, we used satellite-derived data (HOAPS) with a high reputation.
We added "All methods rely on high-quality observational datasets, and the selection of these datasets can influence the results." to the introduction to incorporate the reviewer's concern.

5. **How sensitive are results to the weighting chosen?** (referencing the weights in Table 2)

We tested the dependency of our method on the setting of weights and found reproducible results for varying weights (Geyer, 2026). Figure 7 there (right panel) shows the results of this test and is repeated here for convenience:

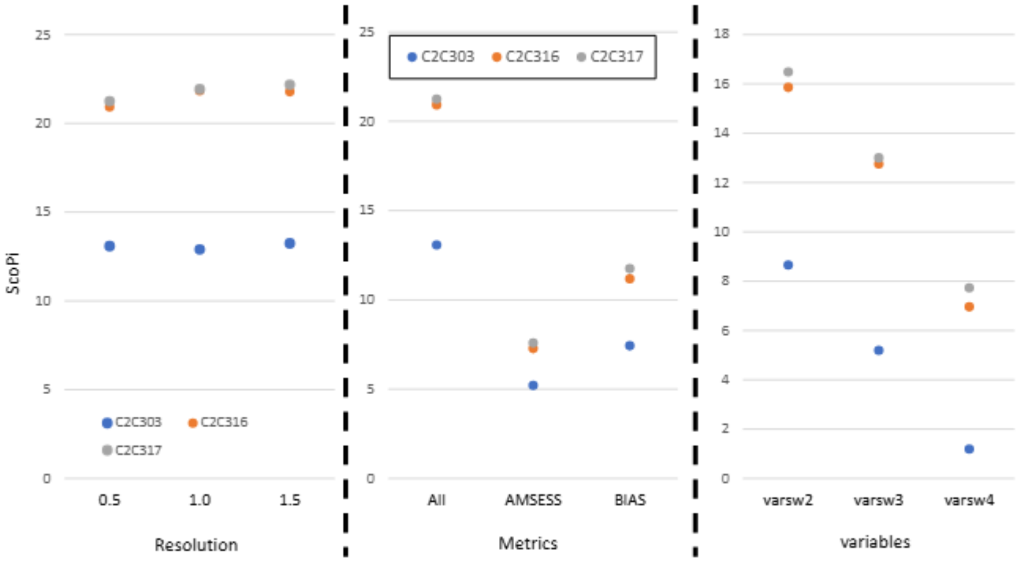

Figure 7. Sensitivity of the $\mathrm{ScoPi}_{simulation}$ with respect to the resolution (left), the metrics (center), and the atmospheric parameters (right). The analysis is done for three model configurations from Phase II (colored points) tested against the reference simulation C2C301 (Tab. 2). 'varsw2', 'varsw3', and 'varsw4' indicate modified weights for the different variables when calculating the ScoPi_{region}. The $\mathrm{ScoPi}_{simulation}$ is calculated by using both weights1 and weights2 from Tab. 5 and then averaging the resulting two scores.

The ScoPi values differ, and the differences between the values are not constant, but the ranking remains, which is the purpose of the ScoPi calculation.

6. The authors note the value of a new score (ScoPi) but only an in-press (I think) publication is provided. This should be laid out in the paper as it seems a key point when scoring across (or accounting for) very different metrics, or at least, key equations and text relevant to this need to be provided in a supplement.

The preprint version of the draft by Geyer (2026) is not yet available although we submitted the revised version on the 14th of March. Therefore we attach the submitted pdf ensure the access for the reviewers.

Nevertheless, we followed the suggestions of both reviewers and included the main equations, see our answer to the major comment 1 of reviewer 1 and Equations 9 to 11 in the manuscript.

7. It was hard to keep track of, but this paper has a lot of **acronyms that may not have always been defined** initially; please check that undefined/unintroduced acronyms are defined before revision.

Thanks for pointing that out. We checked the acronym introduction carefully and added the meanings where necessary.

Addressing some of my general comments above only would add length to the paper, and I do wonder if some parts in the existing manuscript overall can be moved to a supplementary section (?).
Thanks for this suggestion. The longest addition from the submitted to the revised version is the part with the equations for the ScoPi calculation, which were requested by both of the reviewers. In our discussions before the submission, we fought for streamlining everything and leaving out what was not 100% in the scope of the paper. Therefore, we wouldn't move further content to the supplemental information for now.

Specific comments:

Line 3 in abstract: RCM acronym not defined yet, assume it is "regional climate modeling", as introduced in first line.
We added 'regional climate model'.

Line 8: The following text/sentence is confusing to me. Additionally, it is not obvious what a "more efficient parameter combination" is -- can the sentence below be rewritten more clearly? "Comparing the results of optimisation by expert judgement with LiMMo showed that the latter not only confirms the expert judgement focusing on a priori known highly sensitive parameters, but additionally, it allows a model configuration fine-tuning with an explicit control over the tuning process and makes parameter combinations more efficient."
The new version is: "Comparing the results of optimisation by expert judgement with those of LiMMo showed that the latter not only confirmed the expert judgement by focusing on a priori known highly sensitive parameters, but also allowed for fine-tuning of the model configuration with explicit control over the tuning process, making parameter combinations more efficient."
in which 'more efficient' means that small deviations in the parameter values lead to a better consistency of the simulation results with observations. Achieving these small improvements by expert tuning would be very costly in time and computing resources.

Line 13: what does it mean to say the new "configuration could only be reached by using revised datasets" – isn't the new configuration related to optimized parameter combinations?
We changed the wording from 'revised external datasets' to 'updated higher quality external datasets'. Our 'own' contribution to the better model performance is the parameter tuning. But other factors, such as the initial and boundary conditions and the 'external data', influence it as well. External data are the mentioned aerosol distribution, updated soil data map, orography and the vegetation coverage. The related properties of air, soil and roughness influence the energy and moisture transfer and contribute to a more realistic image of reality.

Line 15: The paper suggests an iteration between expert and objective tuning efforts, which to me does not exactly suggest "community-based coordinated parameter tuning" as written in the abstract. What does community-based mean as written here? Is this

the appropriate wording?
With the new strategy of open source modelling with ICON, model development goes faster with more efficient exchange of model developments between the NWP (numerical weather prediction) and the climate community. New model releases are now becoming available twice a year, compared to every 3 years before. Whereas NWP model tuning is driven by the weather services, the usage with climate model setup has to be done by researchers ‘on the side’. Therefore, several hands and minds are needed, and this has to be structured fairly and effectively. ‘Community’ does not mean uneducated model users here.

Line 193: “further fine-tune the” should be “further fine-tuning of the”
Done.

Line 216: are the first two of the 3 parameterizations really not as uncertain? Perhaps the first, but detrained water into an anvil ties more to a time tendency of anvil areas (the purpose of prognostic parameterizations), instead of anvil area (A) itself. No uncertain parameters for these routines?
The cloud cover diagnostics due to convection is not regarded as uncertain, and the parameters of the scheme are not namelist parameters but internal model parameters. For example, the decay time scale of convective anvils is taudecay=1500 s. These internal parameters have not been tuned in this study since there is no ongoing scientific project investigating them. This remains for further studies.

We modified the sentence in the manuscript to “The subgrid-scale cloud cover parametrisation due to turbulence is regarded to be most uncertain and therefore several parameters of the scheme are introduced as tuning parameters in ICON.”

Lines 345-350: it is written: “…in previous studies devoted to objective calibration, monte carlo sampling was used, but they note that complexity associated with a large parameter state space limited the number of parameters to 7 – 8.” This statement about monte carlo efforts is a bit out of date and is beginning to change, especially with use of model ML emulators that enable a tremendous speed increase; see NASA GISS E3 autocalibration work (40 parameters) and MCMC use, as well as discussion in the Carslaw et al. (2026) opinion piece in EGUsphere. High resolution modeling might not quite be ready until emulators of such models are available, which I understand, but monte carlo efforts are definitely expanding into autotuning, but with reliance on AI as model emulators enabling order of magnitude speed-up.

However, our statement concerns exclusively the minimisation of the signal-to-noise error norm function (Eq.7) after the polynomial emulator has already been trained. Earlier works used the Monte Carlo minimum search method (Bellprat et al., 2012, 2016; Avgoustoglou et al, 2022). With several dozen parameters, this method does not allow convergence to the global minimum in practice, unlike the gradient descent proposed in LiMMo (Petrov et al, 2025). These conclusions do not generalise to the advanced AI calibration techniques the reviewer is referring to.

Line 362: “…this minimum may be on the boundary of the constrained region for certain parameters.” This is likely to also happen because of so few parameters too (really the suite is very large, many 10s-100), as well as structural error in the model. How is structural error accounted for? (see above more general point).

We thank the reviewer for this comment and agree that optimised values at the boundary of parameter space indicate a tuning problem. We extended the paragraph accordingly by:

“For the linear regression approximation with a spatial RMSE-based error norm, the target minimisation function ERR(p) is a smooth, convex, scaled Euclidean norm function of the model parameters. This function is known to have only one global minimum. Consequently, the initial parameter values have no impact on the outcome, and the optimisation process always converges to the global minimum. However, this minimum may be on the boundary of the constrained region for certain parameters. This, however, should not be the case for physically consistent parameters and physically meaningful parameter ranges of parameter values admitted.Section~\ref{sec:limmo_tuning} will show that this was not the case for the parameters optimised in this study.”

An answer to the more general question, how to deal with structural error, is given above.

Line 415 and 429 and many other lines: Geyer (submitted) paper – perhaps available now? The authors mention use of a novel metric (Scopi) but there are no details, and the paper cannot be found; information must be provided therefore, so as to demonstrate the great advantage of the new “ScoPi” score as claimed.
The paper is revised and will be available as a preprint soon. But the ScoPi explanations are extended, and the needed equations are added to the text (see answers to the major comment 1 of review 1 above).