Response to reviewer #1 on the manuscript: **Linear Meta-Model optimization for regional climate models (LiMMo version 1.0)**

by Sergei Petrov and Beate Geyer

July 9, 2025

Dear Anonymous reviewer #1, thank you very much for your valuable comments. You will find detailed answers to your comments in the following text. We will refer to the initially submitted version of the manuscript with a red background color and to the revised version with a green background color. We will use a yellow background color for your comments.

Major comment: My main question relates to regularization of LiMMo's regression. LiMMo appears to use linear regression without any sort of regularization, e.g., ridge regression or LASSO regression. A common problem with un-regularized linear regression is that it sometimes yields large "optimal" parameter values that delicately cancel each other's effects. Then those parameter values lead to a poor result when used in a non-linear model like ICON-CLM. However, LiMMo doesn't seem to suffer from this problem in the example tuning run presented (see Fig. 6 and the discussion in the manuscript). Please discuss how this problem is avoided in your run. In addition, please do a tuning run in which the range of each parameter, $p_{\max} - p_{\min}$, is doubled, and then recalculate the R2 values and re-create the plots in Fig. (6). In general, with the range doubled, does regression yield large parameter values that behave poorly in ICON-CLM?

A linear regression model may produce high absolute values for the coefficients of multiple parameters whose effects cancel each other out in the training data. The main reason for this is typically the high correlation between the model parameters. In principle, all physical parameterizations of a well-designed climate model, such as ICON, correspond to distinct physical processes. Of course, the same physical process may be parameterized differently; however the implementation of these parameterizations is always mutually exclusive. In the current study we selected parameters corresponding to distinct physical parameterizations within ICON; therefore, correlation is excluded. The sensitivity of the model quantities to parameter changes is shown in Fig. 5. In the case of multicollinearity, high values of the tendency coefficients (Eq. 7) would lead to extremely high sensitivity values (Fig. 5). This is not the case in our study. Thus, regularization is not essential to the presented manuscript.

Unfortunately, we cannot afford to retrain the linear regression with a doubled range of parameters. There are several reasons for this.

First, reducing the minimum and increasing the maximum limits of the parameter by a factor of two would greatly violate the physical constraints under which the parameterizations were developed. The current minimum and maximum values were selected after extensive discussions with ICON experts and developers. Where feasible, we have already increased the recommended limit values by an additional 10–20% to allow for a wider range of parameter variations.

The second reason is computational constraints. With the 12 continuous model parameters considered in the study, at least 13 additional high-resolution, six-year regional climate simulations of Europe are required. The project's resources are currently close to their limit.

The last point is as follows: Extending the parameter limits for optimization only, without retraining the emulator, will not significantly change the results. The target function (Eq. 4) is a smooth, convex, scaled Euclidean norm function of the model parameters in the case of a linear emulator and a normalized RMSE function

$$\mathrm{ERR}(p_1, ..., p_N) = \sum_n \frac{c_n}{N_t} \sum_k \frac{1}{\sigma_{n,k}} \cdot \frac{1}{\sqrt{N_x \cdot N_y}} \cdot \sqrt{\sum_{i,j} \left(\mathrm{A}_{i,j,k,n} + \sum_m p_m \cdot \mathbf{K}^m_{i,j,k,n} - \mathrm{OBS}_{i,j,k,n} \right)^2}.$$

This is known to have only one global minimum. The only case in which limit parameter values affect the optimization results is when the global minimum for parameter p_m lies outside the range $[p^m_{\min}, p^m_{\max}]$. In this case, the optimization procedure yields either $p^m_{\min}$ or $p^m_{\max}$. In most cases, the optimal parameter values are found within the defined ranges (see Tab. 4). The only limit value obtained is **rlh** for 'equal_weights' and 'tune_temp', so extending the limit values would not significantly affect the LiMMo results.

Minor comment 1: Equation (7): What are the values of Δp_m used in your tuning runs? How is Δp_m related to $p_m^{\min}$ and $p_m^{\max}$?

The specific values of the parameters used to train the linear regression model were selected based on expert knowledge and experience with the ICON model. These values were chosen from within the range $[p_m^{\min}, p_m^{\max}]$. These limit values are listed in Tab. A1 and Tab. A2 in the Appendix. The specific values used for training (tested values) are provided in the Tab. 1:

Parameter	Min value	Max value	Tested values
`tune_albedo_wso(1)`	0.0	0.15	0.0, 0.1
`tune_albedo_wso(2)`	-0.15	0.0	0.0, -0.1
`rlam_heat`	5.0	12.0	6.25, 10.0
`rat_sea`	0.4	1.5	0.4, 0.7
`rat_lam`	0.7	1.3	0.8, 1.0
`rsmin_fac`	0.7	1.5	1.0, 1.2
`cr_bsmin`	80	170	110, 150
`tkhmin`, `tkmmin`	0.2	0.7	0.3, 0.6
`tune_box_liq`	0.04	0.1	0.05, 0.07
`tune_box_liq_asy`	2.5	4.5	3.25, 4
`allow_overcast`	0.8	1.0	0.9, 1.0
`allow_overcast_yc`	0.0	1.5	0.0, 1.0

Table 1: Parameter values used for training of the regression model.

The tested parameter values from Tab. 1 were added to the Tab. A1 and Tab. A2 in the Appendix. The following sentence was also added to the section **The linear Meta-Model approximation**

The specific values of the parameters used for training (tested values) can be found in Tab.A1 and Tab.A2.

Minor comment 2: Lines 320-321: "Test samples were generated by simultaneously varying these parameters within the Latin Hypercube around the minimum and maximum values". Why does this sentence say "around" rather than "between"? Are the samples allowed to include values less than $p_{\min}$ or greater than $p_{\max}$? Can you give more details about how this Latin Hypercube sample is constructed?

Indeed, the formulation "around" is a typo. Thank you for pointing it out. The Latin Hypercube test samples were generated using parameter values between the minimal and maximal values. The sentence

Test samples were generated by simultaneously varying these parameters within the Latin Hypercube around the minimum and maximum values.

is changed to

Test samples were generated by simultaneously varying these parameters from the Latin Hypercube within the intervals from minimum to maximum values.

Minor comment 3: Equation (9): The logical switches, p_l, must take integer values of 0 or 1, but linear regression would seem to yield optimal values of p_l that are real numbers. How does LiMMo convert between the real values yielded by regression and the integer values of, e.g., Fig. (9)?

The logical switches $p_l \cdot \left(\mathrm{MOD}^{p_l=1}_{i,j,k,n} - \mathrm{MOD}^{p_l=0}_{i,j,k,n}\right)$ introduced in Eq.(9) are included in the Meta-Model. However, the logical parameters p_l do not directly participate in gradient descent. These parameters are set to 0 or 1 in advance to simulate the absence or presence of the corresponding switch. This defines only the shift in the linear approximation function. Subsequently, optimization is performed only for the continuous parameters p_m. Considering three logical switches ultimately leads to eight different optimal configurations for each sequence of switches. The final biases for them are shown in Fig.10.

We have added a sentence to the description of the logical switches in section **The linear Meta-Model approximation**. Old version:

... where N_b denotes the number of binary (logical) parameters, and each binary parameter p_l can take the values 0 or 1. The reference simulation assumes $p_l = 0$ for all binary parameters. When $p_l = 0$, the logical switch is off, and no additional signal is added, so the Meta-Model would reproduce the state of the reference simulation. The inclusion of binary parameters introduces constant shifts in the result, but does not affect the gradient of the Meta-Model with respect to continuous parameters.

New version:

... where N_b denotes the number of binary (logical) parameters, and each binary parameter p_l can take the values 0 or 1. The reference simulation assumes $p_l = 0$ for all binary parameters. When $p_l = 0$, the logical switch is off, and no additional signal is added, so the Meta-Model would reproduce the state of the reference simulation. The inclusion of binary parameters introduces constant shifts in the emulator, but does not affect the gradient of the Meta-Model with respect to continuous parameters. Consequently, **minimization involves only continuous parameters, while logical ones are prescribed to 0 or 1**.

Hopefully, it is now clearer that the logical parameters are set in advance and are not part of the gradient descent.

Minor comment 4: Equations (14)-(15): Please clarify the notation "min/max". I was initially confused by whether Eqn.(14) was to be interpreted as really two equations, one for REG_min and one for REG_max, or instead whether REG_min/max was a single variable. It wasn't clear until I reached Eqn.(15) that the former interpretation is the intended one. To clarify, the authors could, for example, simply write the equation before Eqn.(14) as an equation for REG_min and state that a similar equation holds for REG_max.

Indeed, the correct interpretation may not be straightforward. We rewrite the **Sensitivity measure** section as:

To estimate the sensitivity of the ICON-CLM and consequently of the regression model to the considered parameters, the unit-less measure of maximum change $\mathrm{SENS}_{n,m}$ is calculated for each prognostic variable. Firstly we compute the maximal function increments by separately changing all parameters to their limits

$$\Delta\mathrm{REG}^{m,\mathrm{min/max}}_{i,j,k,n} = \mathrm{REG}_{i,j,k,n}\left(p_1^{\mathrm{ref}}, ..., p_m^{\mathrm{min/max}}, ..., p_{N_p}^{\mathrm{ref}}\right) - \mathrm{REG}_{i,j,k,n}\left(p_1^{\mathrm{ref}}, ..., p_m^{\mathrm{ref}}, ..., p_{N_p}^{\mathrm{ref}}\right), \tag{1}$$

where N_p is the total number of parameters, including continuous and logical ones. Here, $\Delta\mathrm{REG}^{m,\mathrm{min}}_{i,j,k,n}$ is the regression increment where only the parameter p_m is changed to its minimum limit. Similarly $\Delta\mathrm{REG}^{m,\mathrm{max}}_{i,j,k,n}$ corresponds to the regression increment when p_m is changed to its maximum. The sensitivity benchmark $\mathrm{SENS}_{n,m}$ of the variable v_n to the parameter p_m is defined as the maximum of the sensitivities revealed for $p_m = p_m^{\mathrm{min}}$ and $p_m = p_m^{\mathrm{max}}$ respectively

$$\mathrm{SENS}_{n,m} = \max\left(\mathrm{SENS}^{\mathrm{min}}_{n,m},\ \mathrm{SENS}^{\mathrm{max}}_{n,m}\right). \tag{2}$$

Eq. 3 gives the expression for calculating the $\mathrm{SENS}^{\mathrm{min}}_{n,m}$ and $\mathrm{SENS}^{\mathrm{max}}_{n,m}$ as the monthly mean signal-to-noise measures (normalized by internal variability $\sigma_{k,n}$) of regression increment where $p_m = p_m^{\mathrm{min}}$ and $p_m = p_m^{\mathrm{max}}$ respectively (Eq. 1)

$$\mathrm{SENS}^{\mathrm{min/max}}_{n,m} = \frac{1}{N_T} \cdot \sum_k \frac{1}{\sigma_{k,n}} \cdot \sqrt{\frac{1}{N_x \cdot N_y} \cdot \sum_{i,j} \left(\Delta\mathrm{REG}^{m,\mathrm{min/max}}_{i,j,k,n}\right)^2}. \tag{3}$$

Minor comment 5: What is plotted in Fig. 6 is not clear to me. I am guessing that Fig. 6 evaluates whether the regression model yields the same result as the ICON-CLM model run for the same configuration and set of parameter values. Is this true? What does each grey dot represent? Is it a single grid point for a single month? Readers might be interested to see plots of other variables, in addition to rsds and pr_amount.

Fig.6 shows a comparison of all grid points and months for the ICON-CLM and the linear regression approximation. The single grey point on the plot represents the monthly value for a specific grid point and validation configuration (i.e., the model parameter values from the Latin Hypercube). The caption of Fig.6 was changed from

Figure 6. The comparison of the regression result (Eq. 6) and the ICON-CLM output for each grid point for training independent setups: (a) monthly mean short-wave radiation flux, (b) monthly sum of precipitation. The dashed red line indicates "perfect match", the value of the R2 determination coefficient is given in the label. Every 100th grid point is shown in the plot.

to

Figure 6. The comparison of the regression result (Eq. 6) and the ICON-CLM output for each grid point for training independent setups: (a) monthly mean short-wave radiation flux, (b) monthly sum of precipitation. Each point on the plot corresponds to specific grid point, month and test configuration from Latin-Hypercube (i.e., all validation cases plotted together). The dashed red line indicates "perfect match", the value of the R2 determination coefficient is given in the label. Every 100th grid point is shown in the plot.

The R2 determination coefficient is one way to assess the quality of the linear emulator. Plots of the variables **tas**, **tasmin**, **tasmax** and **hfls** were omitted from the text because the determination coefficient was high (this was emphasized in L324–325 in the text). These plots can be found in Fig. 1.

However, high values of the determination coefficient can create a misleadingly positive impression of the quality of the linear emulator. In the context of the current manuscript, it is important to compare the RMSE of the difference between ICON-CLM and regression to the measure of ICON-CLM's intrinsic variability. The signal-to-noise ratio is minimized in the end (see the response to thee to the next comment).

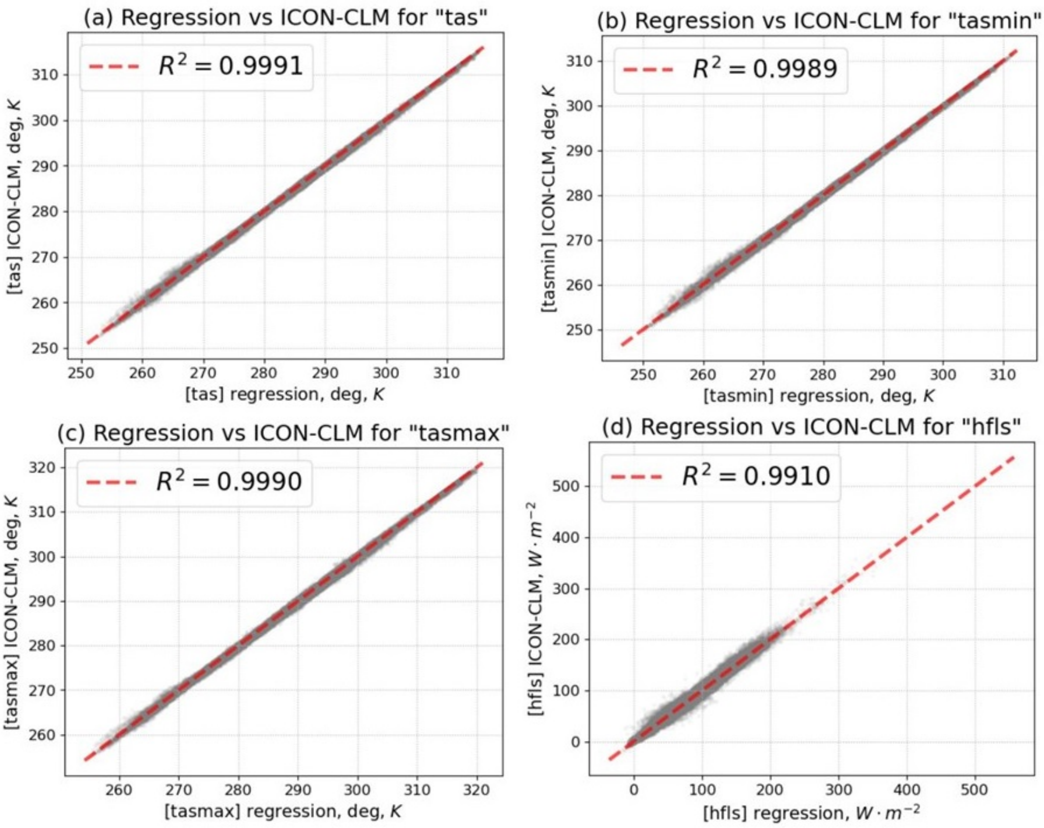

Figure 1: As Fig.6 .Regression vs. ICON-CLM for the variables **tas**, **tasmin**, **tasmax**, and **hfls**. Each grey point shows the monthly mean value of the model quantity for a single grid point in the model domain, for one of the validation configurations.

Minor comment 6: What is plotted in Fig. 7 is also unclear. It is apparently meant to assess the "linear approximation error", but then it plots RMSE relative to obs. However, it's possible for both the regression and ICON-CLM to have the same RMSE but different spatial patterns.

First, I would like to clarify that in Fig.7 (Fig.8 in the revised version), the RMSE to observations of the regression approximation is compared to the RMSE to observations of the ICON-CLM. The regression approximation has not been optimized here; we simply selected the random parameter values from Latin Hypercube, simulated the corresponding setup with the ICON-CLM and compared the results with regression output in terms of RMSE. Therefore, the RMSE of the linear approximation may be larger or smaller than the RMSE of the corresponding ICON-CLM configuration. Section **Meta-Model validation** aims to assess the quality of the linear approximation rather than investigate the quality of optimized configurations; the latter is discussed in the section **Tuning of continuous parameters** .

A different spatial pattern would indicate that the linear approximation error is large, but this large error wouldn't be reflected in RMSE. Also, I don't understand how the ICON-CLM result can sometimes have lower RMSE (better accuracy) than the regression. The regression is approximating the optimum, but often ICON-CLM appears to do even better than the optimum.

This is indeed a very important comment. The initial idea behind Fig.7 (Fig.8 in the revised version) was not only to assess 'linear approximation error', but also to see if the regression approximation could reproduce the order of error norms. For instance, does choosing the linear approximation with the smallest RMSE correspond to selecting the ICON-CLM configuration with the smallest error norm? This holds true for most variables (see Fig.7 or Fig.8).

We agree that a small RMSE does not necessarily imply a perfect spatio-temporal match between the regression and the dynamical simulations. This is a common question in the field of earth system model assessment. However, the current study, as stated in the abstract and introduction, focuses only on an RMS-like measure of quality. We added a new Fig. 2 for a clearer assessment of the approximation. Here, we compare the different sources of error in our analysis numerically. Please note, that we have normalized all the results on intrinsic variability to compare different variables. First (blue) bar shows the typical 'measure of climate model error to observations': RMSE of ICON with NWP configuration to observation. Second (orange) bar shows the measure of intrinsic variability (Eq.2). The third (green) bar shows the measure of 'approximation imprecision': RMSE of ICON to linear regression, averaged over all test cases from Latin Hypercube. We display the temporal average (average for all months) for all quantities.

The results show that the linear approximation error is slightly larger than the intrinsic variability (by a factor of 1.5–1.7) for all variables except for latent heat flux **hfls**. However, the initial error to observations (blue bars) is still much larger than the 'approximation imprecision' (green bars), for the majority of variables. Only for precipitation **pr_amount** the 'approximation imprecision' is only twice smaller than the initial bias. Eventually, this demonstrates the potential for optimization in the LiMMo framework.

We have added the following paragraph to the section **Meta-Model validation**

To assess the inaccuracy of the approximation statistically, we computed the monthly mean values of RMSE between the ICON-CLM output and the linear Meta-Model for each test case in the Latin Hypercube, and plotted the mean values in Fig. 2. As can be seen, the imprecision of the linear approximation (green bars) is slightly greater than the intrinsic uncertainty of the ICON-CLM (orange bars), by a factor of 1.5–1.7 for **tas**, **rsds**, **tasmin**, **tasmax** and **pr_amount**, and by a factor of 2.5 for **hfls**. However, this imprecision (green bars) is still much smaller than the typical error to observations (blue bars) for all variables except precipitation, indicating the potential for optimization.

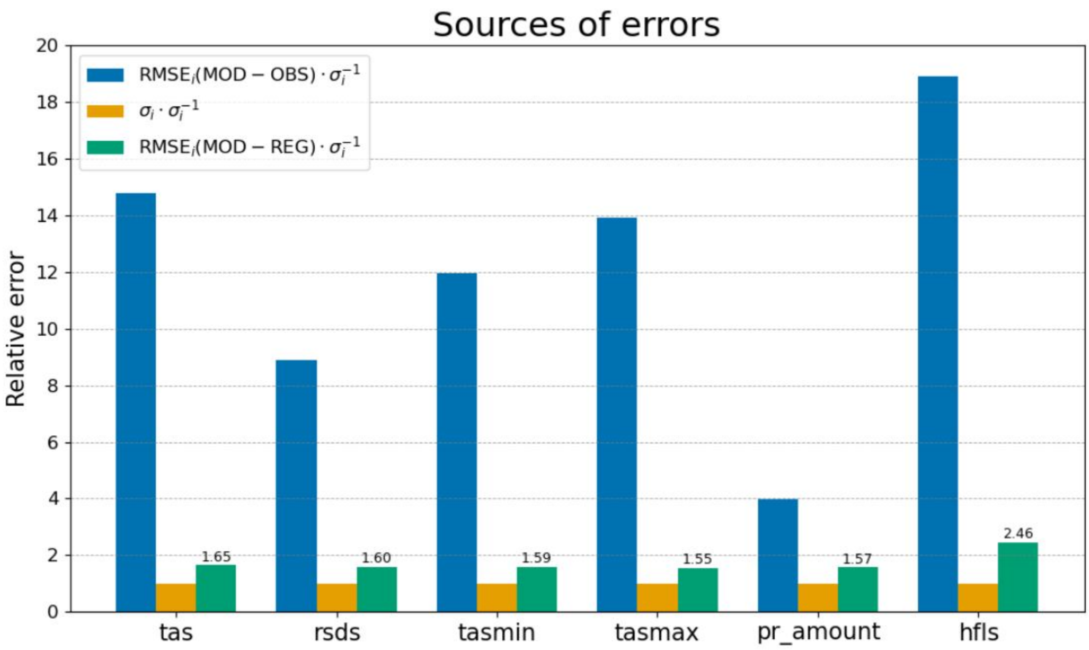

Figure 2: The comparison of the different sources of error in LiMMo. Values are normalized on the intrinsic variability of the ICON-CLM (Eq.2) for each model variable. The blue bar shows the RMSE of the ICON-CLM output with the NWP configuration to the observations. The orange bar shows the intrinsic variability (Eq.2). The green bar shows the RMSE between the ICON-CLM and the linear regression approximation, averaged over all test cases from Latin Hypercube. The temporally averaged values (averaged for all months) are displayed for all quantities.

In addition to Fig. 7, it might be helpful to simply plot spatial maps of, e.g., pr_amount from the regression model next to pr_amount from ICON-CLM.

We have briefly investigated the spatial patterns of 'ICON-CLM - Regression', but these patterns were highly specific to different Latin Hypercube samples. They did not demonstrate common behavior (e.g., overestimation/underestimation in specific European regions), and, therefore, were omitted from the manuscript. Creating an emulator with high accuracy that can reproduce the spatial patterns of a highly nonlinear climate system is very complicated. Training for each model parameter would require dozens to hundreds of dynamical simulations. The LiMMo approach offers an approximation and optimization of the balanced climate state (monthly mean values averaged over a large enough number of years) using the simplest statistical emulator available, making this tuning practically applicable.

In general, we believe that it is not the most accurate approach when a decent configuration is already known. For example, we could only reduce the bias by 10–15% compared to the ICON NWP configuration. However, LiMMo, for instance, could be especially helpful as a quick solution for domains that have rarely been considered before.