Response to Referee #1

We would like to thank the three reviewers for their helpful comments and corrections. We propose a revised version of the paper in which we have taken into account their comments and suggestions. We hope that our revisions will address their expectations.

In the following document, the reviewer's comments are shown in **black** and our point-by-point responses are in **blue**.

The main adjustments proposed in the new version of the manuscript are described below:

- The section "Code and data availability" has been modified to add a brief description of the ICOLMDZ model.
- We have reorganized the figures and tables to improve the flow as reviewers suggested. The figure and table numbers have changed from the initial version of the manuscript:
    - Figure 5 has been removed to avoid repetition with Table 5.
    - Table 3, Table 4, Figure 3 and Figure 8 have been moved to the Supplementary Material.
    - Table 6 has been removed.
    - Figures S2 and S4 were moved from the Supplementary Material to the main text.
    - For each tendency, the results obtained during the first learning and those obtained with the second learning were separated to ease the flow of the presentation of the results. The results of the second learning are only shown when they are discussed (in Section 5 "Refined approach: integrating physical knowledge"). The revised figure numbering, for the zonal wind tendency, is as follows: Figure 4 has been split into Figure 3 and Figure 9, and Figure 6 becomes Figure 4 and Figure 10. We did the same for the five other tendencies in the Supplementary Material. We also separated the results from Table 5 into two tables.
- We propose a new colour scheme for vertical profiles according to the Coblis-Color Blindness:

	Initial training	**Second training with laplacians**
DNN	Green with solid lines	Blue with solid lines
U-Net	Orange with dashed lines	Red with dashed lines

- Section 6 "Discussion" has been merged with Section 7 "Conclusions". The results of this merger is Section 7 "Conclusions and discussion".

- As suggested by the reviewer 3, new results on the emulation of a realistic configuration have been added to the manuscript in Section 6 “Results for a realistic setup”. The data used for this realistic configuration are described in Section 2.1 “ICOLMDZ simulation data”. The terms “for the aquaplanet setup” have been added to the names of Sections 4 “Initial training performance for the aquaplanet setup” and 5 “Refined approach for the aquaplanet setup: integrating physical knowledge” to avoid confusion.

The manuscript "Contribution of physical latent knowledge to the emulation of an atmospheric physics model: a study based on the LMDZ Atmospheric General Circulation Model" studies emulators for all parameterizations of the ICOLMDZ in an idealised aquaplanet configuration. The authors develop four emulators: two FCNs and two UNets, with two sets of input variables: the vertical profile of six state variables and four auxiliary variables in the first set, that are complemented by the vertical laplacian of the state variables in the second set.

Both network architectures are established and have been used to emulate components of GCMs. The main addition by this manuscript is that the authors investigate whether including latent variables as predictors can improve the prediction of physical tendencies. This can be relevant for researchers looking to replace the physical parameterizations with machine-learning based emulators, as stability and accuracy are very important for successful online emulators.

The FCN with vertical laplacians performs similar to the UNet (without vertical laplacians). Considering that the UNet input layer is built up of convolutions that can connect neighboring vertical levels, I find it conceivable that the UNet learns something like a vertical laplacian on the go, without the need to incorporate these predictors explicitly. Incorporating the laplacians also helps with variance, which is underestimated by all emulators.

The methods are outlined clearly and the deep learning code is provided. I believe it would be possible to reproduce the findings, given access to ICOLMDZ and the training dataset. However, ICOLMDZ is not included in the submission.

We have added a link to a repository on the Harvard Dataverse that provides access to the ICOLMDZ model, along with the data shown in this manuscript. The section “Code and data availability” has been modified accordingly.

I would like to ask the authors to address the following points in a revised version of this manuscript:

1. The overall presentation could use some clarification. Due to the nature of the quantities that are evaluated, the figures are quite complex, and the main text discusses findings that are only presented in the supplementary material (S2,

S4). I think it would be beneficial to show the figures that are important for building the main narrative in the paper, and move less important subfigures to the supplemental material.

In the revised version, a few changes have been made to improve clarity.

We have moved Table 3, Table 4, Figure 3 and Figure 8 in the Supplementary Material.

Figures S2 and S4 were moved from the Supplementary Material to the main text.

2. Figures 4 and 5 already contain results that are based on the incorporation of the latent variables that is only introduced in a later section (5.2). I find this confusing, as a reader I was not able to interpret the figures while reading the manuscript front to back. I think it would be beneficial to show results once the underlying experiments have been discussed to keep the flow of information.

   We have taken this comment into account. In the revised version, we have separated the results from the first learning and those from the second learning. Now, the results of the second learning are only shown when they are discussed. The revised figure numbering is as follows:

   - For the zonal wind tendency:
     - Figure 4 becomes Figure 3 and Figure 9.
     - Figure 6 becomes Figure 4 and Figure 10.
   - For the meridional wind tendency:
     - Figure S5 becomes Figure S6 and Figure S8.
     - Figure S6 becomes Figure S7 and Figure S9.
   - For the temperature tendency:
     - Figure S7 becomes Figure S10 and Figure S12.
     - Figure S8 becomes Figure S11 and Figure S13.
   - For the humidity tendency:
     - Figure S9 becomes Figure S14 and Figure S16.
     - Figure S10 becomes Figure S15 and Figure S17.
   - For the liquid water tendency:
     - Figure S11 becomes Figure S18 and Figure S20.
     - Figure S12 becomes Figure S19 and Figure S21.
   - For the ice water tendency:
     - Figure S13 becomes Figure S22 and Figure S24.
     - Figure S14 becomes Figure S23 and Figure S25.

   We have removed Figure 5 as the same information is provided in Table 5.
   We have separated the results from Table 5 into two tables, Table 3 and Table 5.

3. Figure 6 (c-f) contains results that are only introduced later in section 5.2. Again, I find it confusing to see the results early without having the necessary

context, and once the context was there in section 5.2 I had to search for the figure to understand the findings.

We have implemented the required changes to improve clarity, as mentioned just before (see our response to the previous comment).

4. It is obvious that the UNet outperforms the FCN. In the conclusion (Sect. 7), the authors state their interest in coupling both architectures to the dynamical core for an online evaluation. Maybe I am missing a point, but why would it be beneficial to continue working with the FCN if it is outperformed by the UNet? If it is useful to continue with FCN, I would appreciate a short explanation why it should be favored over UNet.

   We have added the following sentences in Section 7 “Conclusions and discussion” to address the reviewer’s comment: “Even though the U-Net architecture outperforms the DNN during offline evaluation, it might not be the case when emulators are coupled with the dynamical core. That is why it would be interesting to assess the performance of both emulators. Moreover, the DNN might be faster to set up than the U-Net since it has six times fewer parameters to optimize”.

5. Please correct the grammar in the first sentence of section 4.2.

   This has been corrected, we have removed the term “study”.

Response to Referee #2

We would like to thank the three reviewers for their helpful comments and corrections. We propose a revised version of the paper in which we have taken into account their comments and suggestions. We hope that our revisions will address their expectations.

In the following document, the reviewer's comments are shown in **black** and our point-by-point responses are in **blue**.

The main adjustments proposed in the new version of the manuscript are described below:

- The section "Code and data availability" has been modified to add a brief description of the ICOLMDZ model.
- We have reorganized the figures and tables to improve the flow as reviewers suggested. The figure and table numbers have changed from the initial version of the manuscript:
    - Figure 5 has been removed to avoid repetition with Table 5.
    - Table 3, Table 4, Figure 3 and Figure 8 have been moved to the Supplementary Material.
    - Table 6 has been removed.
    - Figures S2 and S4 were moved from the Supplementary Material to the main text.
    - For each tendency, the results obtained during the first learning and those obtained with the second learning were separated to ease the flow of the presentation of the results. The results of the second learning are only shown when they are discussed (in Section 5 "Refined approach: integrating physical knowledge"). The revised figure numbering, for the zonal wind tendency, is as follows: Figure 4 has been split into Figure 3 and Figure 9, and Figure 6 becomes Figure 4 and Figure 10. We did the same for the five other tendencies in the Supplementary Material. We also separated the results from Table 5 into two tables.
- We propose a new colour scheme for vertical profiles according to the Coblis-Color Blindness:

	Initial training	**Second training with laplacians**
DNN	Green with solid lines	Blue with solid lines
U-Net	Orange with dashed lines	Red with dashed lines

- Section 6 "Discussion" has been merged with Section 7 "Conclusions". The results of this merger is Section 7 "Conclusions and discussion".

- As suggested by the reviewer 3, new results on the emulation of a realistic configuration have been added to the manuscript in Section 6 “Results for a realistic setup”. The data used for this realistic configuration are described in Section 2.1 “ICOLMDZ simulation data”. The terms “for the aquaplanet setup” have been added to the names of Sections 4 “Initial training performance for the aquaplanet setup” and 5 “Refined approach for the aquaplanet setup: integrating physical knowledge” to avoid confusion.

The manuscript presents the development of data-driven parameterizations for the LMDZ Atmospheric General Circulation Model, specifically in an aquaplanet configuration.

In the first experiment, the authors utilize low-level model variables (e.g., temperature, humidity) to emulate the entire subgrid-scale physics. They train two neural networks (NNs): a simple feedforward neural network (DNN) and a UNet with residual blocks. Both NNs struggle to capture sufficient variance in the output variables. The authors attribute this limitation to an inadequate representation of turbulence, particularly in the boundary layer. To mitigate this issue, they incorporate laplacians of part of the variables among the NN input variables, which may enhance the NNs' ability to represent turbulence. This modification leads to improvements in the R^2 score and the variance of the NN outputs.

The study yields two primary findings:

1. The UNet architecture performs better than the feedforward NN.
2. Incorporating additional variables, thereby embedding physical knowledge into the NNs, can improve their performance.

Overall, the study is comprehensive and well-documented. However, before recommending its publication, I have several suggestions for refinement.

General comments.

- There are many interesting results in this paper. However, sometimes I was lost in the details and between figures and tables. There are repetitions and reminders of previous results, which make the paper hard to follow. For example, Figure 5 and Table 5 ultimately provide the same information. Tables 3 and 4, as well as Figure 3, could possibly be moved to the Supplementary Material. Additionally, it might be useful to merge Table 4 and Figure 3 by including the total number of parameters on Figure 3. Combining Tables 1, 2, and 6 into a single table would also streamline the presentation. These are suggestions, and the authors can take them or not.

  We thank the reviewer for this advice. As suggested, we have:

- removed Figure 5,
- moved Table 3, Table 4 and Figure 3 in the Supplementary Material,
- added the total number of parameters on Figure 3,
- removed Table 6.

- Please proofread. The grammar can be odd in places.

  We have now proof-read the manuscript better.

- Please revise the scale of the colormaps used. The colors are too saturated in Figure 6 (and some of the Figures in the Supplementary) making it difficult to highlight the results.

  In the revised version of the paper, we have applied a logarithmic scale to improve the readability of the results for the cross-sections.

- Your extrapolation regarding execution times might be correct, but in my experience, estimating the gain when using the NN instead of the physical model is much more complex. It also depends strongly on the strategy chosen to implement the NN. I would not risk providing too many details on this subject.

  We totally agree with this comment. Accordingly, we have modified the paragraph concerning the execution time estimates by adding the following sentences: "However, we would like to emphasize that estimating the gain when using an emulator instead of a physical model is much more complex. This is why these estimates should be considered as a preliminary test. It remains to be seen whether this gain would remain when the emulators are coupled with the dynamical core of the model, i.e., in an online setup. For instance, there is generally a substantial overhead when exchanging data from the dynamical core run on CPUs to the physics emulators run on GPUs. One way to alleviate this issue would also be to run the dynamical core on GPUs, which is already possible with the latest development of DYNAMICO".

Point comments.

- Eq. (2): Is this an approximation?

  Yes, it is an approximation. To clarify, we have decided to remove this equation and the corresponding sentences, on lines 134-135.

- line 231: Is it a 1D UNet? I recommend revising the first sentence of the paragraph to state that you used (rather than developed) the UNet architecture.

Yes, it is a 1D U-Net, we are using 1D convolutions. This information is now clearly added to the first sentence of the paragraph. We have also replaced “developed” by “used” as advised in this sentence. Here is the revised sentence: “We have also used a specific architecture, called a U-Net model (Ronneberger et al., 2015), where we use 1D convolutional layers”.

- line 315: Even though I ultimately agree with you on choosing the UNet instead of the DNN, I find that this is not immediately 'obvious'.

  We have removed the terms “it is obvious that”.

- line 380: 'It appears that the U-Net architecture has a better ability to capture the reference tendency than the DNN': wasn't this already the conclusion of Section 4.1?

  We agree, but Section 4.1 focuses on the zonal wind tendency, and Section 4.3 discusses the other tendencies. We propose to modify the sentence by “As was the case with the zonal wind tendency, the U-Net architecture has a better ability to capture the reference tendencies than the DNN”.

- line 465: I don't understand this sentence.

  We propose making some changes to clarify the following sentence “We refer to this laplacian as a hidden variable, often known as a latent variable, because it is obtained directly from the zonal wind”, here is the proposal: “This laplacian is directly inferred from the zonal wind, hence we can refer to this variable as a hidden variable, often called latent variable”.

- line 485: Did the model converge during the 'training with Δ' experiment using the same settings as in the ‘initial training’ experiment?

  Models do not converge perfectly because during training, we use the early stopping criterion to prevent overfitting.This information is now added in Section 3.1 “Neural network architecture” thanks to the following sentence: “It should be noted that this stopping criterion may prevent models from converging perfectly, the key point is to avoid overfitting. Thus, the models saved and used are those that achieved the best performance according to the computed loss metric on the validation dataset tracked during training”.

- line 499-501: Is it a global R^2 score or only for the wind components?

  This section focuses only on the zonal wind tendency: all the scores mentioned correspond to this tendency. Following this comment, we have added “for the zonal wind tendency” to “When comparing the values of the coefficient of determination [...]” to make the text clearer.

Response to Referee #3

We would like to thank the three reviewers for their helpful comments and corrections. We propose a revised version of the paper in which we have taken into account their comments and suggestions. We hope that our revisions will address their expectations.

In the following document, the reviewer's comments are shown in **black** and our point-by-point responses are in **blue**.

The main adjustments proposed in the new version of the manuscript are described below:

- The section "Code and data availability" has been modified to add a brief description of the ICOLMDZ model.
- We have reorganized the figures and tables to improve the flow as reviewers suggested. The figure and table numbers have changed from the initial version of the manuscript:
    - Figure 5 has been removed to avoid repetition with Table 5.
    - Table 3, Table 4, Figure 3 and Figure 8 have been moved to the Supplementary Material.
    - Table 6 has been removed.
    - Figures S2 and S4 were moved from the Supplementary Material to the main text.
    - For each tendency, the results obtained during the first learning and those obtained with the second learning were separated to ease the flow of the presentation of the results. The results of the second learning are only shown when they are discussed (in Section 5 "Refined approach: integrating physical knowledge"). The revised figure numbering, for the zonal wind tendency, is as follows: Figure 4 has been split into Figure 3 and Figure 9, and Figure 6 becomes Figure 4 and Figure 10. We did the same for the five other tendencies in the Supplementary Material. We also separated the results from Table 5 into two tables.
- We propose a new colour scheme for vertical profiles according to the Coblis-Color Blindness:

	Initial training	**Second training with laplacians**
DNN	Green with solid lines	Blue with solid lines
U-Net	Orange with dashed lines	Red with dashed lines

- Section 6 "Discussion" has been merged with Section 7 "Conclusions". The results of this merger is Section 7 "Conclusions and discussion".

- As suggested by the reviewer 3, new results on the emulation of a realistic configuration have been added to the manuscript in Section 6 “Results for a realistic setup”. The data used for this realistic configuration are described in Section 2.1 “ICOLMDZ simulation data”. The terms “for the aquaplanet setup” have been added to the names of Sections 4 “Initial training performance for the aquaplanet setup” and 5 “Refined approach for the aquaplanet setup: integrating physical knowledge” to avoid confusion.

In the manuscript "Contribution of physical latent knowledge to the emulation of an atmospheric physics model: a study based on the LMDZ Atmospheric General Circulation Model", the authors compare two Neural Network (NN) architectures, Dense Neural Networks (DNNs) and U-Nets, for their ability to emulate tendencies produced by a conventional physics package. They find that U-Nets generally outperform DNNs in capturing the variability of these physical processes. The study demonstrates that feature engineering by integrating physical knowledge, such as including the Laplacians of state variables as additional predictors, can improve the accuracy of both investigated NNs, particularly for the DNN.

I think this research could contribute to the development of more skillfull ML emulators of subgrid physics, but it has significant limitations as it is only trained on data produced by an aquaplanet setup without any topography and the performance of the emulators is only evaluated offline. An evaluation of the online performance, coupled to the aquaplanet ICOLMDZ setup would certainly be more meaningful than the offline evaluation.

We have taken this comment into account and added new results related to the emulation of a realistic configuration in Section 6 “Results for a realistic setup”. The data used are presented in Section 2.1 “ICOLMDZ simulation data”. The abstract, the introduction and the conclusion have been modified accordingly.

Some of the findings presented in this manuscript appear to be preliminary and would benefit from more thorough evaluation. I will elaborate on this in the following points:

Major comments

1. From my perspective the biggest issue, as mentioned, is that the developed emulators are only evaluated offline. A small stability/performance test of the DNNs/U-Nets coupled online would be very insightful.

   We totally agree that online evaluation is crucial. However, we would like to point out that it is out of the scope for this first study because the main objective of this article is to show that adding predictors with physical meaning helps to obtain an emulator closer to the physical model. Before studying emulators in an online setup, we need to ensure that they reproduce the behaviour of the physics in offline mode. Moreover, online evaluation requires procedures and metrics that differ from offline evaluation and an entire paper would be required to examine this thoroughly. Thus, the online evaluation is presented as a perspective in Section 7 “Conclusions and discussion”.

2. Line 160: Why don't you consider 2D outputs such as precipitation. I am sure the physics package of ICOLMDZ also parameterizes some 2D fields.

   Our goal was to reproduce the physical tendencies used in the calculations, and thus, we did not consider 2D outputs such as precipitation as mentioned because these variables are diagnostic. This information is now added in Section 2.2 “Input and output variables of emulators” of the revised manuscript.

3. Line 198: Why do you choose one every 20 cells and not a random subset? With this fixed "step size" you potentially overfit to these locations, don't you? What about evaluation on full data set, it is mentioned here, but you never come back to that as far as I saw?

   We have chosen one cell every 20 cells because the data is well distributed across the globe, all latitudes are represented, as illustrated in Figure 1 below. This information is now added to the paragraph: “In this way, we obtain data that is well distributed across the globe and we ensure that we have sampled at all latitudes (see Fig. S1 in the Supplementary Material)”. The revised version also includes the figure below.

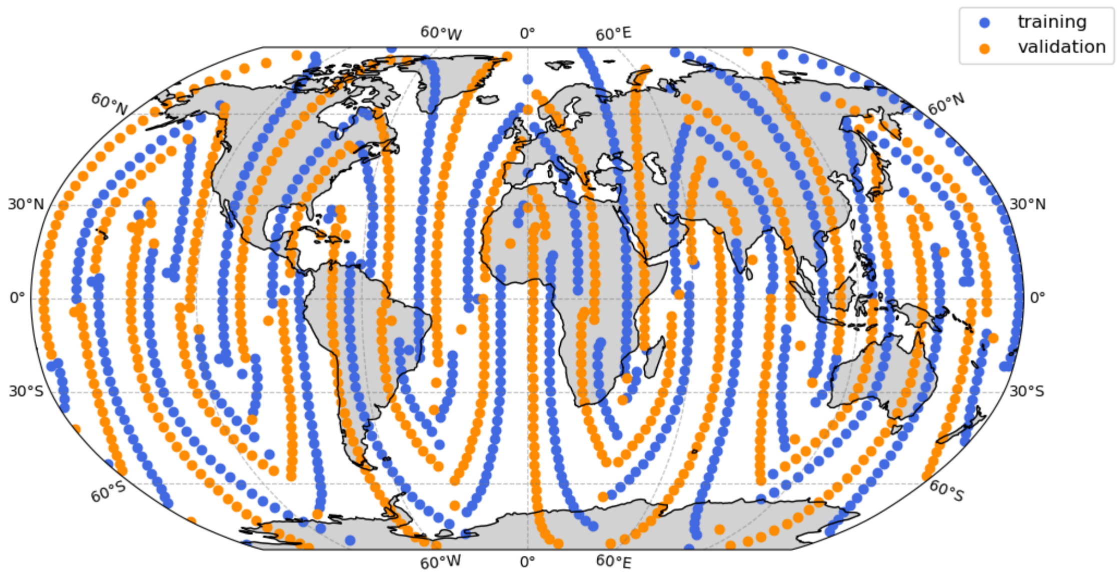

Figure 1. Distribution of grid cells over the globe used for training in blue and validation in orange, for the realistic configuration. The same distribution of grid cells is used for the aquaplanet.

All the results presented in the manuscript come from the test dataset where only temporal sampling is performed.

4. Line 224: You only tried this particular setup (6 layers and 512 neurons in each layer) and no hyperparameter optimization at all?

   We tried many different setups for the DNN (and also for the U-Net architecture) and we decided to work with a DNN composed of 6 layers of 512 neurons for each layer since it appeared to be a good trade-off between the computation time and the quality of the results. This information is now clearly added to the paragraph.

   We did not use hyperparameter optimization techniques but we did some manual tuning.

5. Line 483: To judge impact of adding additional predictors adding a plot learning/validation curves to assess convergence would be helpful. Also, best to train multiple models to see if the improvement is "statistically significant"

   We agree that plots corresponding to the loss function for the learning and for the validation step help to assess convergence of the models. However, we have not added this type of plots to the manuscript because the interpretation is challenging due to the multivariate nature of the emulation problem.

   Models do not converge perfectly during training because of the early stopping criterion. This information is now added in Section 3.1 “Neural network architecture”: “It should be noted that this stopping criterion may prevent models from converging perfectly, the key point is to avoid overfitting. Thus, the models saved and used are those that achieved the best performance according to the computed loss metric on the validation dataset tracked during training”.

   For instance, the training and validation loss curves for the DNN without laplacian and for the DNN with laplacians are shown below with Figures 2 and 3, respectively. The early stopping criterion used to prevent overfitting, with a patience of 20, must be taken into account when reading the loss curves. The model selected for the DNN without laplacian achieved a validation loss of 5.0×10^{-1}. For the DNN with laplacians, this score is equal to 4.2×10^{-1}.

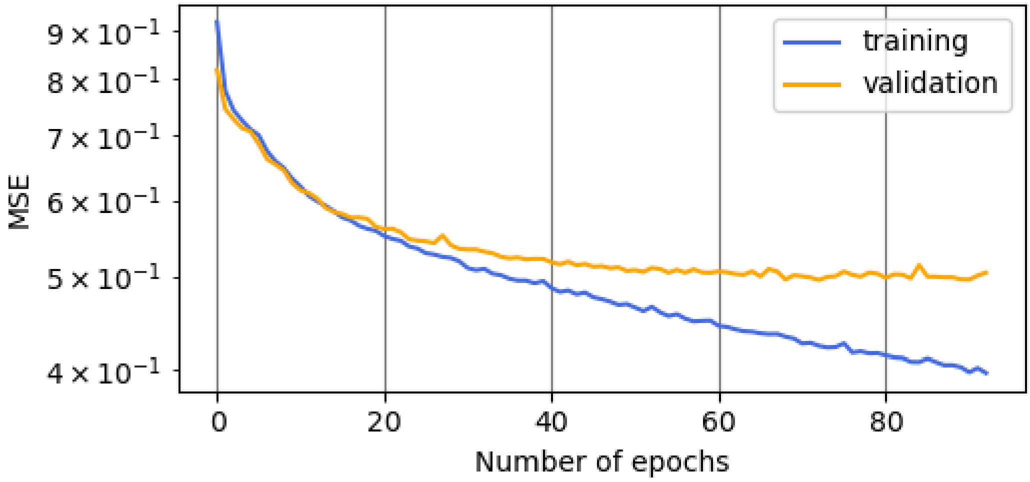

Figure 2. Plot of training and validation loss over epochs for the DNN without laplacian.

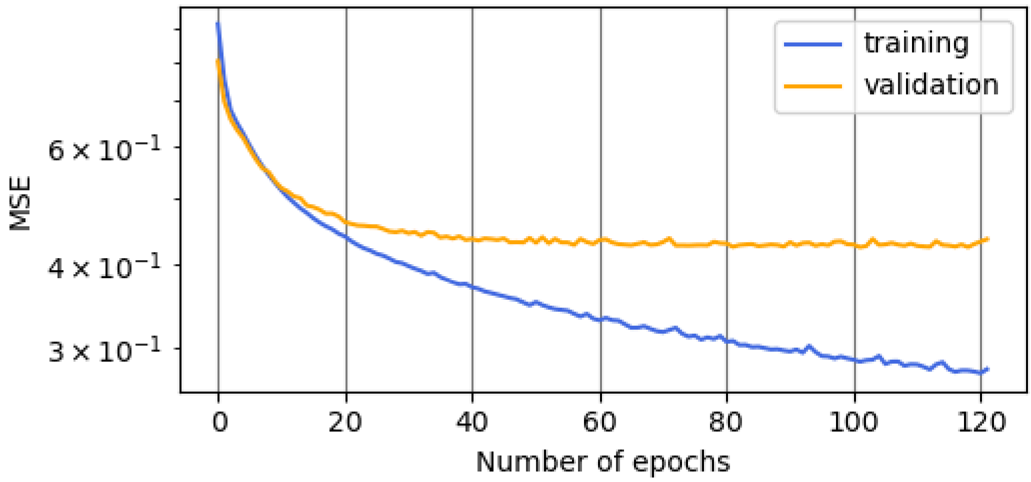

Figure 3. Plot of training and validation loss over epochs for the DNN with laplacians.

We have trained each model several times but we have not kept all the corresponding loss values. However, the results are as follows:

- Similar behaviours and performances in terms of loss function for a chosen model have been observed.
- The laplacian-aware emulators achieve a lower loss than the initial emulators without laplacian.
- U-Net models perform better than DNN models in terms of loss minimization.

The following sentence has been added to the manuscript: “Each model has been trained a few times, but as we observed similar behaviours and performance in terms of loss function for a chosen model, we only kept one of each”.

6. Line 562: The discussion is very short and not complete, you should add some limitation of you study, e.g, the aquaplanet setup.

   We have merged Section 6 “Discussion” and Section 7 “Conclusions” to create a new Section 7 “Conclusions and discussion”.

   Several discussion points are now provided regarding some elements to understand why the U-Net is more efficient than the DNN, the contribution of the turbulence process for the five other tendencies, and also the execution time estimates.

   We are also discussing some perspectives that can be considered as direct limitations of the study, such as the capacity of the emulators to generalize and the online evaluation.

   We did not include a paragraph in which we discuss the use of an aquaplanet as a limitation because we have added some results, in Section 6 “Results for a realistic setup”, on the emulation of a realistic configuration, even though these are preliminary results.

Minor comments

1. Eq. (2): I doubt that the ICOLMDZ uses as simple Euler step for time integration.

   Yes, it is an approximation. To clarify, we have decided to remove this equation and the corresponding sentences (lines 134-135).

2. Line 242: Please provide more details how you "reformatted" the input to 2D tensors. You are using 1D convolutions as previous developed ML physics schemes using U-Nets, right?

   We have taken this comment into account and we have revised the paragraph accordingly. Here is the information that has been added: “Data is reformatted before being provided to the encoder. For vectors, but also the three scalars spread across all vertical levels, we concatenate each atmospheric column in order to obtain a 2D matrix of dimension 79 multiplied by the number of 1D variables. This data format is compatible with the 1D convolution operator that processes each atmospheric column independently”.

   Yes, we are using 1D convolutions. This information is now clearly added to the paragraph: “We have also used a specific architecture, called a U-Net model (Ronneberger et al., 2015), where we use 1D convolutional layers”.

3. Eq. (8-13): I don't think explaining the computations of means, variances, and metrics in such detail is necessary, but that is ultimately your choice.

   We have decided to keep these computations in the manuscript.

4. Line 420/Figure 8: Why do you choose a random timestep and not multiple ones and take a median/mean profile? Also, it could be helpful to plot A,B, and C in the same figure for easier comparison.

   We thought it would be interesting to have a look at results for some random timesteps before we got an overview of the breakdown. For this reason, we present the results of the breakdown for one timestep in Figures S1, 8 and S3 (now these figures are all provided in the Supplementary Material). The mean vertical profiles, obtained from all vertical profiles over one year, are presented in Figures S2, 9 and S4 (now, these figures are all included in the main text of the manuscript).

   If we plot A, B and C in the same figure, it will not be easy to see anything because the scale of the zonal wind tendency differs depending on the point considered: 10^{-4} for A, 10^{-3} for B and 10^{-5} for C, for one time step for instance.

5. Line 423: What physical process is represented by the thermals opposed to convection and turbulence?

   To respond to this comment and the following comment, we have replaced "These physical processes include turbulence d_u_turb, convection d_u_conv, thermals d_u_therm, gravity wave drags associated to fronts d_u_gwd_front and those associated to precipitations d_u_gwd_precip" with the next sentence "These physical processes include turbulence d_u_turb, boundary-layer convection d_u_therm, and convection modelled as static adjustment d_u_conv, gravity wave drags due to emission by fronts d_u_gwd_front and those due to emission by convective precipitations d_u_gwd_precip". More details on these processes are given in Hourdin et al. 2020 (3.1. Small-Scale Turbulence, 3.2. Boundary Layer Convection, 3.3. Deep Convection, 3.6. Radiation, 3.8. Gravity Waves).

6. Line 436: "This is also the case for gravity wave drags from precipitations" - Do you mean convectively coupled waves, or what should "gravity waves from precipitation" mean?

   Please, see our response to the previous comment.

7. Line 470: Is delta z really fixed and if so, what is this fixed distance?

   δz is fixed and is equal to 1. It corresponds to a difference between levels (and not to a difference in altitude). We have modified the text and here is the revised version we propose: "In the denominator, delta z corresponds to the difference between vertical levels. We consider $\delta z = 1$. This distance is fixed as the step between vertical levels is constant".

8. Line 613: You should mention the CPU inference time of your models as well and mention that even if you are able to use the GPU for inference during coupling, you have an overhead from copying the data from CPU to GPU as long as ICOLMDZ is not running on the GPU as well.

   We have modified the paragraph concerning the execution time estimates by adding the following sentences: “However, we would like to emphasize that estimating the gain when using an emulator instead of a physical model is much more complex. This is why these estimates should be considered as a preliminary test. It remains to be seen whether this gain would remain when the emulators are coupled with the dynamical core of the model, i.e., in an online setup. For instance, there is generally a substantial overhead when exchanging data from the dynamical core run on CPUs to the physics emulators run on GPUs. One way to alleviate this issue would also be to run the dynamical core on GPUs, which is already possible with the latest development of DYNAMICO”.

9. Line 631: Watt-Meyer et al., 2024 did not emulate a physics package but derived the subgrid physics by coarse graining from high-resolution simulations

   Thanks for pointing this out, we apologize for the confusion. We have removed the quote in question.

Typos/Grammar

1. Line 19: The second "numerical" is redundant

   Fixed.

2. Line 201-203: "These numbers of samples are multiplied by the length of the input vector X′ and the output vector Y′, for all the three subsets, which correspond to variables aggregated over the vertical dimension" What do you mean by multiplied by length of ..., to get the number overall size of the input data?

   Exactly, we added this sentence to illustrate the size of the data. However, in the new version, we have decided to remove this sentence to improve clarity.

3. Line 229: "It is trained with the learning rate scheduler of 10^{−3} ..." What exact scheduler did you use? I assume the factor 10^{-3} indicates some decay or did you mean learning rate of 10^{-3}?

   Our apologies, it is a mistake. The results presented in this manuscript come from models with a fixed learning rate, we did not use a scheduler here. Therefore, we have removed the term “scheduler”.

4. Line 230: I wouldn't say that "a U-Net is a CNN architecture" but that it uses convolutional layers

   The initial sentence has been changed by “We have also used a specific architecture, called a U-Net model (Ronneberger et al., 2015), where we use 1D convolutional layers”.

5. Line 246: "Once the models had been built, we trained them." Not necessary to state this.

   Noted, we have removed this sentence.

6. Line 257: "In this article" - You mean section I think

   We have replaced “In this article” with “In the following sections”.

7. Line 284: "We will sometimes compute the RMSE on specific subsets of the test dataset to exhibit performances that vary for instance across latitudes or vertical levels." - "exhibit" does not fit here

   Corrected, we have replaced “exhibit” with “assess”.

8. Line 333: "..., we examine them study by latitude belts, ..."

   This has been corrected, we have removed the term “study”.

9. Line 416: "We decided to carry out this study on three grid cells..." - please use columns instead of grid cells here

   This has been corrected as suggested by the reviewer.

10. Line 429-431: please rewrite the sentence

    The initial sentence was “Turbulence movements (see Figure S1a, 8a and S3a) are present whatever the point studied near the surface but also at higher altitudes, between vertical levels 20 and 40, for points located at high latitudes and in the subtropics”. We have changed this sentence and here is the new proposal: “Whatever the point considered, turbulent movements (see Figure S3b, S4b and S5b) can be observed near the surface. These movements also occur at higher altitudes, more precisely between vertical levels 20 and 40, for the points located at high latitudes and in the subtropics”.

11. Line 535: reformulate "with these new learning"

    Done, it has been replaced by “with the laplacian-aware emulators”.

12. The word "Indeed" is overused throughout the manuscript.

We have revised the text to remove several occurrences of “indeed” (9 terms in total now, compared to 18 initially).

Reference

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., et al. (2020). LMDZ6A: The atmospheric component of the IPSL climate model with improved and better tuned physics. Journal of Advances in Modeling Earth Systems, 12, e2019MS001892. https://doi.org/10.1029/2019MS001892