Improvement and calibration of clouds in models

Tuning and Sensitivity Session

April 13, 2021, Météo-France, Toulouse, France

Feel free to ask your questions through this document during or after the different talks. The objective is to be able to extend discussion as there is no real coffee break.

1- A machine learning assisted stochastic cloud population model as a parameterization of cumulus convection

Samson Hagos, Jingyi Chen, Katelyn Barber, Koichi Sakaguchui, Zhe Feng, Heng Xiao, Robert S. Plant

A machine learning assisted cloud population model is coupled with the Advanced Research Weather Research and Forecasting (WRF) to represent fluctuations in cloud base mass flux associated with the life cycles of cumulus convection cells. In this model, the size distribution and associated cloud base mass flux of the convective cells are related to their previous state and the change in the convective area via a transition function. The convective area tendency in turn is assumed to depend on cloud base mass flux resolved by the host model (WRF). The transition function is represented by a neural network trained by a 1 km grid spacing WRF simulation. The cloud population model continuously predicts the cell size and cloud base mass flux distributions which are then fed to an entraining plume model. It is shown that such an approach to parameterization can lead to realistic precipitation statistics and diurnal cycle over various regions as well as MJO related propagation of precipitation.

Question 1 <Doris Folini>: (WE CAN KEEP THE QUESTION IN THE DOC, I MAY HAVE MISSED IT IN THE TALK) Could you use observations instead of / in addition to a high resolution model to train for / learn about f(c) ?
<Samson Hagos> Yes it can be done to a degree. For example in this work we use observation only. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS001798 But one needs mass flux to do the kind of parameterization that we are building which we have no observation for.

Question 2 <Frédéric Hourdin>: (WE CAN KEEP MY QUESTION IN THE DOC) Do you have constraints on your f0 f1, b0 b1 ? Are your running a mass flux model at the end ? I may have missed the equations
<Samson Hagos> Yes the machine learning model only predicts the pdf of mass flux. A random sub-set of the pdf is then fed to a cloud model. Frédéric : And so do you have some constraints that you should guarantee on those variables ?
Question 3 - Fabian Senf <senf@tropos.de>: How do you define "cells" and "cell area"? I guess this involves segmentation of a field which can have various sensitivities. And as follow-up: You mentioned cell interactions: Do you have a measure how cells interact with each other?
<Samson Hagos>In this earlier work we use the machine learning model offline to examine interactions. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS001798

Question 4 - Hannah Christensen: To what extent are the improvements you see (e.g. MJO) due to the stochastic model in general versus the ML learned transition function
<Samson Hagos> We have not thought about that for we have not untangled the stochasticity from the machine learning based pdf from which we are sampling.
Question 5 - Catherine Rio: does that mean that once you have the cloud-base mass-flux right, your scheme is doing everything else quite ok (mixing, precip, …)?

<Samson Hagos> The cloud model is fairly standard and seems to work OK. In addition to cloud base mass flux, the cloud model is fed cell sizes which affect entrainment. We were careful to make sure we use a run-of-the mill mass flux model to isolate the impact of the cloud population model.

2-Development of a physically-based parameterization of the raindrop formation processes through machine-learning

Azusa Takeishi, Chien Wang

While quick transformations of cloud droplets into raindrops are frequently observed, reproducing such fast conversions in atmospheric models has been one of the major challenges. Recently some studies showed the importance of turbulence in expediting the cloud-droplets-to-raindrop process. In addition, it has long been argued that aerosol size distributions and chemical compositions may be critical to the collision-coalescence process. In this study, we utilize a parcel model that calculates the processes of aerosol activation, condensational growth, and collision-coalescence, all based on physical equations. We apply machine-learning algorithms to model-simulated raindrop mass alongside ten dynamical and microphysical variables as input features. Differences between the machine-learned results and those predicted by empirical parameterizations, as well as the applicability of the machine-learning-based parameterization in a regional model, will be discussed.

Question 1 <Fleur Couvreux>: I may have misunderstood but you mention Sulfates and sea particles Is your parcel model also suitable for other types of aerosols?

Answer by Azusa Takeishi: Yes, certainly. The initial size distributions of aerosols can be specified by users, with any hygroscopicity values & size distribution parameters, for how many aerosol species/populations one would like. But for my machine-learning procedure, I used sulfate & sea salt populations. Their size distributions vary (selected by Latin-hypercube sampling) within a defined range; their number concentrations vary too, so there are some simulations with very few sea salt (or sulfate particles) as well.

Question 2 <Frédéric Hourdin>: Is it not a case where we could be afraid of not having a physical relationship between inputs and dQr/dt ? with the risk of being completely wrong outside the learning domain ? When scanning climate conditions from equator to pole and from surface to stratosphere ? Welcome :)

Answer by Azusa Takeishi: The inputs (or features) are 53 variables, including T, P, S, w, eddies, droplet/rain size distributions (separated into tens of bins each). Even if there is no physical connection between some of the 53 variables and dQr/dt, the machine will be able to learn that; at the end we are able to know which input variables are most important by ranking the "importances". So, having some variables that are not connected to dQr/dt (or ones with the least importances) as input features is not a problem (we can eliminate those "extra" inputs to increase the machine-learning efficiency at the end). If we feed the machine with a set of input features that are outside of the learning range, then the machine would not be able to make a good prediction - in order to minimize this possibility, we train the machine with more than a million samples.

Question 3 <Danny Williamson>: Why Random Forrest? Do you think anything is lost by not having predictions with uncertainty? And finally, how accurate would prediction have to be to replace a physical parameterisation (you have 0.66 accuracy right now, how high does it need to be)? I guess that was 3 questions :)

Answer by Azusa Takeishi: We're currently using the ensemble learning methods (Random Forest & Gradient Boosting) because some tests with the Multi-Layer Perceptron & the Support Vector regressors did not work well (haven't explored other methods yet). I'm not entirely sure if I know enough about making predictions with uncertainty, but I will look into it right away! Today I didn't talk about the scores/accuracy because there may have been some errors in the calculation of the scores and the actual scores might be much higher than 0.66 (cannot check it now because the supercomputer is on maintenance). I think scores higher than 0.85 would be "good enough" but

this is just an arbitrary number in my mind at the moment. I should see other machine-learning studies for reference.
Follow up (DW): Also see statistical models :). Random Forest is attractive because it can be quite accurate for low effort, but perhaps you might look to moving to a deep NN or similar, particularly if you can run on the supercomputer… :)
Question 4 <Axel Seifert>: Have you tried solving the resulting ODEs? Estimating the process rate (autoconversion) is only the first step. Solving the ODEs can reveal problems that you may not have seen in the process rates (https://doi.org/10.1029/2020MS002301).
Answer by Azusa Takeishi: Yes, ODEs are solved in the Pyrcel model. But thank you for the reference - I'll read it right away and add more answers asap.
Question 5 <Maxime Colin>: Thanks for the talk. Why comparing your tests with the Kessler scheme only?
Answer by Azusa Takeishi: I'll be comparing my results with a few more parameterizations other than the Kessler scheme too, very soon. The Kessler scheme was plotted out because that was the simplest & just for a reference.

3 - Assessing calibration issues and intrinsic limits of a convection parameterization using machine learning

Romain Roehrig1, Daniel Williamson2,3, Fleur Couvreux1, Frederic Hourdin4

Over the past ten years, the CNRM climate model development team has developed and implemented a convective scheme that aims at representing dry, shallow and deep convection in a unified and continuous way. During its integration, the question of its effective ability to capture the various convective regimes arose. The present work investigates this latter question in a rigorous manner: does a set of tuning parameters exist for this parameterization? Or are we facing its intrinsic limits? This investigation is made possible thanks to (i) the recent availability in our SCM of a wide range of single-column model convective cases, each of them being associated with relevant references (the new input standards for SCM forcing shared within the French community were instrumental for this achievement), and (ii) the development of machine-learning tools that provide an objective framework for process-based model tuning, and which ultimately illuminate the issue of parameterization tuning.

Question 1 <F. cheruy>: Would it be interesting to redo ARM cu case with interactive radiation and coupling with surface (including a LES reference)?
Answer (F Couvreux): There is for example the new “ARM” case proposed by Y Zhang et al (2017) that can include interactive surface and radiation and that has been built on a composite of 15-year of observations. This case should soon be in the DEPHY format.
Further answer (Romain): Yes, especially since the work of N. Villefranque within the HighTune project now provides relevant reference for radiative metrics. For coupling with the surface, yes also, but it is needed to get forcing parameters for the surface. Probably in an idealized framework, with consistent LES.

Question 2 <Doris Folini>:Several questions, feel free to reply only in the google doc. How do you choose your 24 parameters? Expert knowledge or all parameters that are there? How important is this choice? Can you tell a posterior that 10 parameters would have been sufficient? And use this info in the future?
Answer (DW). The question is, are the 10 parameters always the most important? It can be that 10 parameters are crucial over the initially explored ranges, but that as you reduce the space with history matching, you zoom into a region where new parameters are dominant. If you never varied the other parameters, you don’t see this.The approach treats the parameters that are “inactive” as noise in the emulators, until such time as they are activated (if they ever are).
Further answer (Romain): To say it another way, the dominant parameters that explain the sensitivity of the SCM may evolve across the waves, and one parameter may be rapidly “fixed” within a small range. It is hard to know from the start I think, and at least the framework provides a rigorous way to confirm our initial intuitions (or not).

Question 3 <Chris Holloway>: How do you know that your “best” choice of parameters doesn’t have a “cancellation of errors” problem? Or is there another approach used to try to determine which parameterisations and parameter values are most physically relevant on their own before being tested in combination with the other parameterisations and parameter choices? OK thanks!
Answer (Romain): I agree that there may be some compensations between two parameterisations. An example as I mentioned is between the turbulence and mass-flux parameterisations. One way is to add metrics that may help disentangle such balance, but only if we have “good” reference for that, with small uncertainty (otherwise, it won’t constrain anything). Another way is to consider cases which involve only one parameterisation (e. g., our work in Audouin et al. 2021 JAMES, where we could only consider the turbulence parameterisation). But not sure if it is always possible, especially because, for some regimes, the interactions between parameterisations is often as important as the parameterisations themselves. Besides, my “best”

choice was only for illustration. Ideally, all “best” choices should be considered, and at least their diversity should be documented (and this can help to find new metrics that we may be able to further constrain, if relevant).

Question 4 <Maxime Colin>: Very interesting presentation! Are you able to understand in which cases the 1D and 3D biases are the same, and in which cases they are different?
Answer (Romain): Thanks Maxime. For several 3D systematic errors, we find them in relevant 1D cases, especially for boundary layers, low clouds. It’s also the case for the diurnal cycle of deep convection. For deep convection regimes, it’s a bit more difficult, probably because the interaction with the large-scale dynamics is also key (WTG/DGW may be an interesting alternative to further use, with however the question of relevant references). For the CINDY2011/DYNAMO case, we found large similarities with the 3D model in the Indian Ocean for the temperature and specific profiles. Besides, even though the biases are different between 1D and 3D, from my experience, we often see similar sensitivity to parameterisation parameters. But it’s harder to use that in a tuning framework, which involves a priori many parameters...

4--A network approach to the geometric structure of shallow cloud fields

Franziska Glassmeier1, Graham Feingold2

The representation of shallow clouds and their radiative impact is one of the largest challenges for global climate models. While the bulk properties of cloud fields, including effects of organization, are a very active area of research, the potential of the geometric arrangement of cloud fields for the development of new parameterizations has hardly been explored. Self-organized patterns are particularly evident in the cellular structure of Stratocumulus (Sc) clouds so readily visible in satellite imagery. Inspired by similar patterns in biology and physics, we approach pattern formation in Sc fields from the perspective of natural cellular networks. As an outlook, we discuss how a similar network approach can be applied to describe and quantify the geometric structure of shallow cumulus cloud fields.

*Talk has been withdrawn

5-Machine-learned atmospheric optical properties for radiative transfer computations

Menno Veerman1, Robert Pincus2,3, Caspar van Leeuwen4, Damian Podareanu4, Robin Stoffer1, Chiel van Heerwaarden1

A fast and accurate treatment of radiation is essential for high quality atmospheric simulations, but radiative transfer solvers are computationally very expensive. In this study, we use machine learning to determine the optical properties of the atmosphere, which control the absorption, scattering and emission of radiation. We train multiple neural networks to predict all optical properties calculated by the RRTM for General circulation model applications - Parallel (RRTMGP) from the pressure, temperature, water vapour content and ozone mixing ratio of each grid cell. The predicted optical properties are highly accurate and resulting downward longwave and shortwave radiative fluxes have errors within 0.5 W/m2 at the surface. Depending on the size of the neural networks, our implementation is up to 4 times faster than RRTMGP. We thus conclude that neural networks are very suitable to emulate the calculation of optical properties.

Question 1 <Axel Seifert>: Usually the split is in training data, validation data and testing data. I did not see a validation set being used. Also the training plot (score as function of EPOCHs) showed only one line, but should show 2 lines for training and validation set. In addition, having a small testing set (only 5 %) can be misleading if those 5 % are drawn from an easier part of the phase space.
Ok, thanks, this is confusing, but now it makes sense. Thanks for clarifying. Nevertheless, you should show both lines in the EPOCH time series. The 5 % is maybe fine, if your training set is very large compared to the number of trainable parameters (which is probably the case for your small ANNs). Nevertheless, using an easy validation or testing set, can be a problem.
Reply <Menno Veerman>: I see that it is somewhat confusing. Generally, I did not see striking differences between the training and evaluation loss (hence why I decided to only show the evaluation loss to reduce the number of lines in the plot).
Axel: That training and validation loss collapse onto a single line is good and something worth mentioning.

Question 2 <Sophia Schäfer>: Did you look into how errors depend on region, cloud type, meteorological situation,...? Thanks, I missed that it was clear-sky only. Do you plan to also look at cloudy situations? I focussed on the gaseous optical properties because they are much more expensive to compute than the cloud optical properties, I expect much less of a speed-up for the cloud optical properties. Therefore, I was not directly planning to look into the cloud optics as well, but a full optical properties emulator might be a nice followup indeed.

Question 3 <Doris Folini>: Once you have run your algorithm, can you tell which g-points are more relevant for / contribute more to the error? If so, could such information improve our overall understanding of what matters more / most to get things right? What gaseous species, for example, matter more / most where on the planet or during what time of the day / year?

Answer: Although I mainly focussed on the broadband fluxes, I have looked at the fluxes per band (RRTMGP has 14 SW bands of 16 g-points and 16 LW bands of 16 g-points). Mean absolute errors per band are generally between 10^-3 and 10^-1 W/m2, but up to 0.2 W/m2 for the smaller networks I tested. I have not really looked at the errors of the separate g-points. I don't think you can learn from this approach which gaseous species are most important where and when, merely which g-points respond more/less linearly to changes in gas concentrations, and are thus easier/harder to predict with a neural network approach

Question 4 (Maxime Colin): Why did you decide to use neural networks only to model the relationship between atmospheric properties and optical properties, and not for the second step to compute radiative fluxes? Is it just because this is where more improvement was needed?

6-Parameterization and tuning of cloud and precipitation overlap in LMDz

L Touze-Peiffer, F Hourdin, C Rio

We will present the effect of a new parameterization, inspired by Jakob & Klein (2000), to take into account cloud overlap in the formation and evaporation of precipitation in the atmospheric model LMDz. In this parameterization, at each level, the precipitation flux is divided into clear-sky and cloudy precipitation fluxes, with corresponding precipitation fractions. Newly formed precipitation falls within the clouds, thereby increasing the cloudy precipitation flux. By definition, the cloudy precipitation falls in saturated air, so only the clear-sky precipitation flux evaporates. This simple treatment corrects some inconsistencies found in the standard version of LMDz and leads to a better representation of cloud and precipitation overlap.

The version of LMDz containing this new parameterization is tested and tuned using the HighTune tools over various 1D case studies and in 3D simulations. The results illustrate the potential of machine learning techniques to guide the development of new parameterizations.

Question 1 <Fleur Couvreux>:Which metrics did you use for the tuning? was surface rain used?
So surface rainfall not used.

Answer: Indeed, we used the same metrics as in Hourdin et al. (2020), so no metrics on surface rainfall. Nevertheless, I investigated quite systematically the best simulations of the standard version of LMDz obtained after tuning, and no simulations predicted realistic rain rates in ARMCU, RICO and SANDU. While not a proof, it suggests that the tuning of realistic model clouds in the standard version of LMDz is not compatible with realistic surface rain rates (contrary to the new version). To prove it rigorously, I should indeed use metrics of surface rain rate in addition with the metrics already used.

Question 2 <Etienne Vignon>:Is your last conclusion about Kuhn' concept application not related to the fact that a model is not reality (by definition)? I mean, a model cannot be exhaustive in terms of process representation.

Answer: A model is indeed not the reality and cannot be exhaustive in terms of process representation, but a climate model attempts to predict accurately certain aspects of the reality. We could make a distinction between *intermediate variables*, that are used in parameterizations but should not be predicted as accurately as possible (ex: mass flux in convection schemes), and *target variables* of the model, which the model should get right (ex: cloud fractions). When analyzing model results, we only look at model target variables. The difficulty consists in finding an appropriate balance between the realism of the different target variables of the model.

Question 3<Daniel Williamson>: It is clear, whatever we get right in 3D, we do so for the wrong reasons. So if progress is degraded, but for good reasons, is that necessarily bad? *Follow up:* So can we use bad (by which I mean right for the wrong reasons) models to predict climate change? (Conceptually. Of course, I understand we must do something)

Answer (Ludo): Good question! The answer depends on what we call ``for the wrong reasons" I guess. The errors on intermediate variables seem to me completely acceptable (see above). The errors on target variables of the model are more questionable. However, I would say that for the prediction of climate, there is a hierarchy of target variables: if we do not want to get a totally unrealistic climate, some global properties must be guaranteed, for instance to get a correct radiative balance at the top of the atmosphere (see Hourdin et al. 2017). We might accept some errors in other target variables in order to get these more important target variables right.

DW Follow up again… Top of the atmosphere radiative balance is an interesting example. Typically we don't get it "right" but we get it "right for the coupled model" in order to get reasonable 20th century warming. This usually means the TOA we tune to is not the observed. Whether we view this necessary step as a result of accumulation of errors in the processes, or something else, it is interesting to note that what we accept to be "right" and what we accept to be "good reasons" or "acceptable compromises" are largely a result of the history of development for that model and for the community in general.

(Ludo) I fully agree… For me, one potential of High-Tune explorer is to make these different choices previously hidden in the development of models explicit at the level of the scientific community and thus opens a debate on what is "acceptable" or not, what we accept to be "good reasons" and which tuning targets are the most relevant.

Let's note also that at the end there is an operational need to make models work, that is to produce realistic results… even if the method used to make it work is not fully justified.

Question 4 (Danahé Paquin-Ricard): Can you learn from the tuning parameters proposed by High-tune about the interactions between the parameterization? What's the physical meaning of the proposed tuning? Is it possible to gain some insight on the model with this methodology?

Answer (Ludo): from my (limited) experience, definitely. Using the High-Tune explorer is actually much more efficient to explore the interactions between the parameterizations. Without it, there is a risk of being blocked in a local optimum in the region of parameters. On the contrary, as High-Tune explores a much larger sample of parameter space, it helps to understand what a climate model can and cannot do. Expert judgment is still crucial to analyze the results of simulations and understand the structural limitations of a climate model, but High-Tune guides this expert judgment rather than replace it.

Question 5 (Maxime Colin): Great talk, Ludo! With your cloud evaporation change, what was improved and what was degraded (after re-tuning)? Do you have ideas to address the different points that were degraded (if that's relevant)?

Answer (Ludo): Thanks a lot Maxime. The answer was in the slides that I did not have time to show. It seems that it is possible to have a shortwave cloud radiative effect at the top of the atmosphere better than the one given by the standard version of LMDz. Nevertheless, whereas we expected also significant improvements in terms of surface precipitation, the results were more disappointing. The results were, at best, of similar quality to those given by the standard version of

LMDz, and in the Tropics, they seemed to be slightly degraded. The analysis of results for other metrics is still ongoing.

7-Characterising Convection Schemes Using Their Linearised Responses to Convective Tendency Perturbations

Yi-Ling Hwong1, Siwon Song1, Steven Sherwood1, Alison Stirling2, Catherine Rio3, Romain Roehrig3, Chimene Daleu4, Robert Plant4, David Fuchs1, Penelope Maher5, Ludovic Touzé-Peiffer6

We study the behaviour of convection schemes by probing their responses to small perturbations in temperature and moisture tendencies following the linear response function method of Kuang (2010). We compare 12 physical packages in 5 atmospheric models using single-column model (SCM) simulations under RCE conditions. Results are also compared to that of a cloud-resolving model (CRM). We show that the procedure is able to isolate the behaviour of a convection scheme from other physics schemes. We identify similarities but also substantial differences among the SCMs and between the SCMs and the CRM, some of which can be explained by scheme physics. All SCMs display kinks in their responses, which are absent from the CRM, suggesting that they might be related to switches or thresholds embedded in convective parameterisation. The models' moisture responses are related to their RCE profile, while their temperature responses do not and therefore can be regarded as independent diagnostics.

Question 1 <name>:
Question 2 <name>:

Q3 <Maxime Colin>: Why does the CNRM scheme behave similarly to adjustment schemes?
<Answer YLH> We are not really sure, perhaps Romain or someone from the PCMT team can provide a hypothesis. One thing that sets the PCMT scheme apart from the other mass flux schemes in this study is its consistent use of buoyancy as the forcing term in its formulation (triggering, entrainment/detrainment, mass flux calculation etc), and it could be that this "consistency" improved the responsiveness of the scheme to local perturbations
<Romain> I don't have much intuition on that yet… But this is clearly something that we need to investigate.

Q4 <Roel Neggers> Really interesting results! I was wondering if the same signatures of the convection schemes are visible in GCM output.. so not in SCMs, which of course are a bit constrained in how parameterized convection can respond to a perturbation. Did you look into this? We studied this for marine low level clouds for a different intercomparison study: https://doi.org/10.1002/2015MS000503

<Answer YLH> Thanks! That's a good question. We have not directly looked into if these signatures are also visible in GCM output. In fact - our results seem to have illuminated areas that were NOT visible in 3D simulations, e.g. the flaw in LMDZ6A (overestimation of evaporation in shallow cloud layer) was not previously identified in 3D studies. And I think there are a few studies that showed that the Kain Fritsch and BMJ schemes produce similar results for extreme rainfalls, but in our experiments their responses are vastly different. So in summary, I don't know, but I suspect the LRF method probably highlights signatures that are not obvious in GCM outputs, hence provides a complementary view/method to evaluate convective parameterisation. Thanks for the paper! I think I have actually read your paper, just need to recall (and reread) now to see what you found :-)

8- Sensitivity of observable output of modeled microphysics to stochastically perturbed parameters

Tomislava Vukicevic1, Derek Posselt2, Aleksa Stankovic1

This study investigates sensitivity of cloud and precipitation parameterized microphysics to stochastic representation of parameter uncertainty to evaluate a potential to use satellite remote sensing to estimate properties of such uncertainty representation for the purpose of ensemble modeling. A stochastically perturbed parameterization scheme (SPP) is applied to multiple microphysical parameters within a lagrangian column model. The experiments based on ensemble simulations indicate a high sensitivity of simulated microphysics-sensitive satellite observables to the SPP properties. The analyses suggest a strong potential to use the satellite observations to estimate the modeled stochastic uncertainty of the parameterizations. Further analyses will be presented at the conference.

Question 1 <name>:
Question 2 <name>:

9-A Hybrid Deep Learning CNN-LSTM Model for Predicting Monthly Rainfall Over Japan

Paul Adigun

Predicting rainfall is an important task due to the high reliance on it, especially in the agriculture sector. Prediction is difficult and even more complex due to the dynamic nature of rainfall. In this study, we carry out monthly rainfall prediction over some selected locations in Japan. The rainfall data were obtained from the Japan Meteorology Agency (JMA). We study the predictive capability of a hybrid Convolution Neural Network (CNN) integrated with Long Short Term Memory (LSTM) model on the parameters recorded by the automatic weather station in the selected regions. The developed hybrid model is applied to four different locations in different climatic regimes in terms of monthly precipitation characteristics. Overall, this study was able to establish the skill performance of hybrid CNN-LSTM model in capturing complex relationship between the causal variables and monthly variation of rainfall in the study region.

*Talk has been withdrawn

10-Tropical free-tropospheric humidity differences in global storm-resolving models

Theresa Lang1, Ann Kristin Naumann2, Bjorn Stevens2, Stefan Alexander Buehler1

Tropical free-tropospheric humidity (FTH) plays a key role in controlling the Earth's outgoing longwave radiation (OLR), but it is poorly simulated by conventional climate models. We investigate whether global storm-resolving models (GSRMs) simulate FTH more accurately, by quantifying inter-model differences in the multi-model ensemble DYAMOND. We find that the model spread in FTH is approximately halved compared to conventional climate models. However, the differences still cause a considerable spread of 1.2 Wm2 in tropical mean clear-sky OLR. To reduce this spread a reduction of humidity biases would be most beneficial in the lower and mid free troposphere. In the horizontal, FTH biases in two moisture regimes have a particularly strong impact: Dry subsidence regimes and moist regions adjacent to deep convection. In the most critical regions FTH biases are related to parameterized processes (e.g. microphysics, turbulence), so a better understanding of how they affect FTH is needed.

Question 1 <Doris Folini>: Have you also looked whether the larger spread humidity in CMIP5 is due to “all” models or whether some are decidedly better / worse than others and contribute more to the spread? As many people (still) work with these models, such info may be of interest to people.

Answer (Theresa Lang): Overall, I do not think that the larger spread in CMIP5 is only due to a few single models. But there are a few models that stick out because they do not even reproduce the typical C-shape of the RH profile. Unfortunately I cannot name these models now, but I can check and maybe add this information here later.

Question 2 Hannah Christensen: Really interesting work - many thanks! How well sampled is variability in FTH in the 40-day DYAMOND runs compared to many years of CMIP data? Are the DYAMOND runs more similar because they don’t have time to sufficiently drift away from the starting state, or do the common SSTs in both constrain the solutions?

Answer (Theresa Lang): Thanks, these are important questions. To your first question: Of course the inter comparison is limited somewhat by the shortness of the runs. However, we are confident that the differences we see are really systematic model biases rather than the result of a poor sampling of natural variability. We estimated natural variability from several years of ERA5 data (this provides an upper bound for the variability one could expect in DYAMOND, because of the varying SSTs). The DYAMOND inter-model spread is significantly larger than the inter-annual variability in ERA5. To your second question: I forgot to mention that we only use the last 30 days of the simulations for the comparison. The first 10 days are excluded to minimise constraints due to the common initialisation. Of course we cannot rule out that the DYAMOND models are still in their transition from the initial conditions even after the first 10 days. The model spread does not systematically increase over the 30 day period we analyse, which makes us confident that the models are already in their own equilibrium. However, we will not be able to make a final statement until longer model integrations become possible.

Question 3 by Azusa Takeishi: Thank you for your talk - I don’t know too much about DYAMOND, but does the better representation of humidity mean the better representations of clouds in DYAMOND as compared to CMIP5?

Answer (Theresa Lang): This is an interesting question. From what I have seen there are still substantial differences in cloud properties in DYAMOND, so I would be careful to claim that a better representation of humidity improves the representation of clouds. Unfortunately I have not looked at clouds in detail, so I can’t say more about this. But there are people looking at cloud properties in DYAMOND, e.g. Karol Corko (DLR).

11-Microphysical sensitivities in global storm-resolving simulations (SRMs)

Ann Kristin Naumann

In global SRMs that resolve convection explicitly instead of parameterizing it, microphysical processes are now fundamentally linked to their controlling factors, i.e., the circulation. While in conventional climate models the convective parameterization is one of the main sources of uncertainties (and a popular tuning parameter), this role might be passed on to the microphysical parameterization in global SRMs. In this study, we use a global SRM with a one-moment microphysics parameterization and do several sensitivity runs, where in each run we vary one parameter of the microphysics scheme in its range of uncertainty. First results indicate that microphysical sensitivities in global SRMs are substantial and resemble inter-model differences such as in the DYAMOND ensemble. Among the parameters tested, the scheme is particularly sensitive to the ice fall speed and the width of the raindrop size distribution, which both cause several 10s W/m2 variation in radiative fluxes.

Question 1 <Frédéric Hourdin>: a few W/m2 globally means a few K on global mean sruface temperature in a coupled model. Question : is the very different cutoff for large IWP important for cliamte ? Understood ?
<Ann Kristin Naumann>Thinking of Theresa’s clear-sky radiative transfer calculations, I think, it can be. But this is also a region that is cloudy, so cloud radiative effects might mask radiative effects of water vapor, at least partly.

Question 2 <Chiel van Heerwaarden>: How big is the uncertainty in ice fall speeds? What is the way to validate this? Can observations help us or are these too hard to make?
<Ann Kristin Naumann>Ice fall speed is a popular tuning parameter. We increased the fall speed by about a factor of 3 to the original value of Heymsfield and Donner (1990). I am not familiar with direct measurements of the ice fall speed but I could imagine some more indirect ways using observed cloud cover or cloud radiative effects.
Comment on Q2 <Axel Seifert>: As far as I understand, this is bulk fall speed. Hence, it includes a lot of unknowns like particle size distribution, particle habit, etc. This supports the factor 3 estimate. Fall speed alone would have a smaller but still considerable uncertainty (for a known particle size and habit).

Question 3 <Tomi Vukicevic> have you considered perturbing multiple parameters? Sensitivity may change depending on parameter coupling
<Ann Kristin Naumann>I agree. As an example: Because in many ways the “narrower RSD” and the “higher ice fall speed” simulations show opposite effects, we were interested to see what happens if we change both parameters in the same simulation. And indeed the two effects do not cancel but we see, e.g., a substantial decrease in high cloud cover and a higher increase in net TOA fluxes.

12- Impact of middle atmospheric humidity on boundary layer turbulence and clouds

Neelam Namadev Malap1,2, Thara Prabhakaran1, Anandakumar Karipot2

We investigated the moisture present above the boundary layer and its association with the cloud development and boundary layer(BL) turbulence using Large Eddy Simulation (LES). The dry boundary layer with shallow clouds observed during Cloud-Aerosol Interaction and Precipitation Enhancement Experiment (CAIPEEX) over the arid Indian peninsula is studied here. LES derived fluxes and variances are compared with the constant altitude aircraft observations at different elevations. LES simulations with the drier conditions above the BL resulted in deeper, warmer and drier BL with an enhanced boundary layer turbulence. Drying has resulted in energetic BL eddies and a doubling of moisture exchange coefficients. LES sensitivity indicates that the middle atmospheric water vapor alone could influence the shallow to deep cumulus cloud transitions in the monsoon regime in a dramatic way. A 30% drying above the BL could drastically reduce the liquid water path and cloud albedo by 10-15%

Question 1 <Chris Holloway>: Thanks, interesting talk! Can you explain a little more how you performed the drying perturbation in the LES sensitivity study? Also, can you please give the reference for the publication that you mentioned? Thank you.

<neelam> Thank you Chris. We have introduce drying above 1km into water vapor mixing ratio by subtracting water vapor with 10% of default water vapor, 20% default water vapor and so on.

here is the link for the paper- https://doi.org/10.1016/j.jastp.2021.105553

Question 2 <Fleur Couvreux>:Can you also please detail the rad experiment? that were shown on the figures?

<neelam>Those are radiation simulations. We just switched on the radiation parameterization i.e. RRTM for both long wave and short wave into WRF LES.

13-Interpretable Machine Learning and Remote Sensing for Cloud Detection and Classification

Thomas Chen

The detection and classification of atmospheric clouds, in both ground-based and satellite-based domains, is crucial to the understanding of Earth's energy balance, climate, and weather. Utilizing the latest deep learning methodologies, we train a convolutional neural network, of the ResNet50 architecture, on cloud imagery data from the ground. Building off of previous work, we explore the interpretability of models, i.e. how and why the models are coming to their decisions. This requires ablation studies concerning different input modalities and the fine-tuning of particular features like the learning rate and loss functions. We also propose the generation of gradient class-activation maps for the visual representation of the inner workings of our machine learning models. These figures, which act as saliency maps, allow us to break open any black box models and understand why they are making their decisions. Interpretability is key to be aware of biases within the models.

*Talk has been withdrawn

INVITED TALK:

Climate model development and the role of machine learning

D Williamson, Wenzhe Xu, Bertrand Nortier

Machine learning and AI, thanks to a number of high-profile successes, are often touted as potentially holding the keys to solutions to some of science's greatest challenges. Indeed Google's DeepMind, having achieved some of the aforementioned successes, turned their attention to weather and climate prediction. But what is machine learning and can it replace climate modelling (my view is no)? In this talk I will introduce some key ideas in machine learning, and argue how it can assist in climate model development, for example through tuning and falsification of parameterisations.

In the last half of my talk I will focus on recent innovations with Gaussian processes and their implications for tuning. In particular I will introduce 2 ideas: Kernel History Matching (KHM) and Integrating Gaussian Processes (IGPs). KHM is a new tuning method capable of searching parameter space for climate models with important emergent features without requiring that these features occur in exactly the same place or at the same times as they do in reference data sets (knowing that in general they can't due to limited resolution). IGPs use (deep) Gaussian processes to couple components of a network (e.g. a climate model). I will demonstrate the technology and look ahead to its implications for tuning.

Question 2 <Maxime Colin>: What's your definition of a kernel?
Ans: A kernel is just a function k(x,y): X x Y → R. We typically deal with positive definite kernels. Kernels are nice because defining a kernel for a space gives an inner product to that space (and there is an equivalent unique Hilbert space associated with that kernel).

Question 3 <Martin Köhler>: How can a ML technique emulate a on/off process such as saturation adjustment when the weights are linear? (clouds, convection are all very non-linear)
Thanks, right, it's the non-linear transfer functions. (MK)

Question 4 <Etienne Vignon>: Is the orthogonality assumption behind the EOFs somehow limit to capture complex patterns?
Not at all. The limit is in fact the richness of the ensemble. So there are only n degrees of freedom in a training set and, if n < N(gridcells) you don't have a full basis for the whole space. This is true with the kernel method as well. The question is, how rich is the response to parameter changes. If it is near chaotic, there is probably no hope, but I don't think we think that.
Thank you! (EV)

Question 5 < Fleur Couvreux> : what do you need as a training dataset to emulate the different gaussian processes in your last result which propose several layers of emulators if I understood correctly?
You need the models that will be coupled.

Question 6 <Tomi Vukicevic> : Given that weather and ES models solve initial and boundary value problem and solution also depends on a large number of control parameters, all leading to a very large number of degrees of freedom in AI/ML parameter space, could pre selection be avoided in an objective way usin ML tools?
I'm not exactly sure what pre-selection is. Is it a formal method?
Pre-selection is referring to the need to reduce the number of control parameters for developing an emulator. Are there ML tools to address that? NWP and ES models have of the order of 10**10 controls including IC and BC
Ah I see. So ML has formal methods for sensitivity analysis etc that picks out active parameters and reduces the degrees of freedom etc. Whether you would just pass a whole NWP model into a machine learning algorithm is kind of questionable (I think process-based methods are more stable and break the problem down somewhat). The basis methods I talked about are what I tend to use. So the fact you have 10**10 controls really means you have a number of spatio-temporal fields. Basis methods project these forcings onto a basis so that you end up with far fewer controls (the coefficients of basis vectors) and then you can form the control fields from there. Hopefully this partially answers the question, but happy to take follow ups by email.

Question 7 <Nicolas Rochetin> : Can we say that ML used to bypass parameterization is just meant to mimic the flow, when a parameterization tries to represent the flow ? Would this semantic suggestion sound reasonable ? A representation includes an underlying conceptual picture, which is absent from ML used to simulate the flow.
DW: I think it's more than a semantic suggestion and it certainly is reasonable. ML is physics free and hence constraint free. But if you have a statistical model that represents a flow really well, it would seem like a great idea to study it to see how it physically handles the representation.

Discussion:

High-Tune:Explorer tutorial

Follow-up session on Friday 10 am at:
https://gather.town/app/dounBToM8YBQ0e5I/labLMDZ

Put your name here if you would be interested in having a late session on Friday evening French time (normal day time in the West)

-
-
-

Ask your questions or comments here!

-
-
-