DARPA ACTM: Milestone 3 Report

AIBEDO: A hybrid AI framework to capture the effects of cloud properties on global circulation and regional climate patterns

Prepared by Kalai Ramea*
with inputs from Hansi Alice Singh†, Soo Kyung Kim*, Peetak Mitra*, Subhashis Hazarika*, Dipti Hingmire† and Haruki Hirasawa†

*Palo Alto Research Center, Inc.
† University of Victoria

April 13, 2022

The main objective of this milestone is to identify appropriate assessment metrics for the hybrid model and deliver prepared datasets that can be used for Phase 1 model development. In addition, this report includes updates on hybrid model development, outcomes of the initial model runs, and the methodology plan for upcoming modules. Our official documentation page is here: https://aibedo.readthedocs.io.

Overall Update

1. Preprocessed 100% of the Phase 1 Earth System Model output datasets. Metadata details are provided in this report, and the data files can be shared upon request.
2. Identified additional input variables to improve our model performance based on our *alpha* model results. Those variables are also preprocessed and included in the dataset.
3. Developed appropriate performance metrics to assess accuracy and speed of the model.
4. Produced initial results of our updated model and have elucidated plans for the next steps.
5. Updated our official documentation page to include details about the hybrid model and climate dynamics between cloud and atmospheric circulation.

Datasets

Our training data for Phase 1 consists of a subset of CMIP6 Earth System Model (ESM) outputs which had sufficient data availability on AWS to calculate the requisite input variables for our analysis (shown in Table 1). For each ESM, there are three sets of data hyper-cubes: (a) input, (b) output, and (c) data for enforcing physics constraints. Based on the initial results from our alpha hybrid model, we revised and increased the list of input variables to achieve better hybrid model performance. The updated list of input, output, and constraint variables is shown in Table 2.

Table 1: Earth System Model datasets for Phase 1 training

Model	N historical	N ssp585	lat spacing (°)	lon spacing (°)
CESM2	1	0	0.942	1.2500
CESM2-FV2	1	0	1.895	2.5000
CESM2-WACCM	1	0	0.942	1.2500
CESM2-WACCM-FV2	1	0	1.895	2.5000
CMCC-CM2-SR5	1	0	0.942	1.2500
CanESM5	5	0	2.789	2.8125
E3SM-1-1	1	0	1.000	1.0000
E3SM-1-1-ECA	1	0	1.000	1.0000
FGOALS-g3	2	1	2.278	2.0000
GFDL-CM4	1	1	1.000	1.2500
GFDL-ESM4	1	1	1.000	1.2500
GISS-E2-1-H	1	0	2.000	2.5000
MIROC-ES2L	3	0	2.789	2.8125
MIROC6	1	0	1.400	1.4062
MPI-ESM-1-2-HAM	1	0	1.865	1.8750
MPI-ESM1-2-HR	1	0	0.935	0.9375
MPI-ESM1-2-LR	1	0	1.865	1.8750
MRI-ESM2-0	1	0	1.121	1.1250
SAM0-UNICON	1	0	0.942	1.2500

Table 2: Variable list and descriptions

Data Type	Variable Name	Description
Input	clwvi	Mass of cloud liquid water in a column (kg/m^2)
Input	clivi	Mass of cloud ice water in a column (kg/m^2)
Input	cres	TOA Cloud radiative effect in shortwave (W/m^2) (rsutcs - rsut)
Input	cresSurf	Surface Cloud radiative effect in shortwave (W/m^2) (rsds - rsus - rsdscs + rsuscs)
Input	crel	TOA Cloud radiative effect in longwave (W/m^2) (rlutcs - rlut)
Input	crelSurf	Surface Cloud radiative effect in longwave (W/m^2) (rlds - rldscs)
Input	netTOA	Net TOA radiation (all-sky) (W/m^2) (rsdt - rsut - rlut)
Input	netTOAcs	Net TOA radiation without clouds (clear-sky) (W/m^2) (rsdt - rsutcs - rlutcs)
Input	netSurf	Net Surface radiation (W/m^2) (rsds - rsus + rlds - rlus - hfss - hfls)
Input	netSurfcs	Net Clearsky Surface radiation (W/m^2) (rsdscs - rsuscs + rldscs - rlus - hfss - hfls)
Input	lcloud	Cloud fraction averaged between 1000hPa and 700hPa (%)
Output	tas	2-metre air temperature (K)
Output	psl	Sea Level pressure (Pa)
Output	pr	Precipitation ($kgm^{-2}s^{-1}$)
Constraint	ps	Surface Pressure (Pa)
Constraint	evspsbl	Evaporation ($kgm^{-2}s^{-1}$)
Constraint	heatconv	Convergence of vertically integrates heat flux (Wm^2)

The ESM data are pooled together to form the training and testing datasets for our hybrid model. However, it is important to note there are substantial differences in the climatologies and variability of

some of the chosen input variables across models (Fig. 1). In particular, global average cloud liquid water content, cloud ice water content, and net top of atmosphere radiation vary more across ESMs than other variables. The former two are the result of differences in cloud parameterizations between ESMs, while the latter is likely due to uncertainties in the overall magnitude of anthropogenic forcing over the historical period. Comparing spatial correlation scores (Fig. 2), shows net TOA radiation fields are very similar across models while the spatial pattern of cloud ice and water content varies substantially. Such variations represent the inter-ESM uncertainty in the representation of the climate. However, many of these ESM differences are largely removed during preprocessing described below.

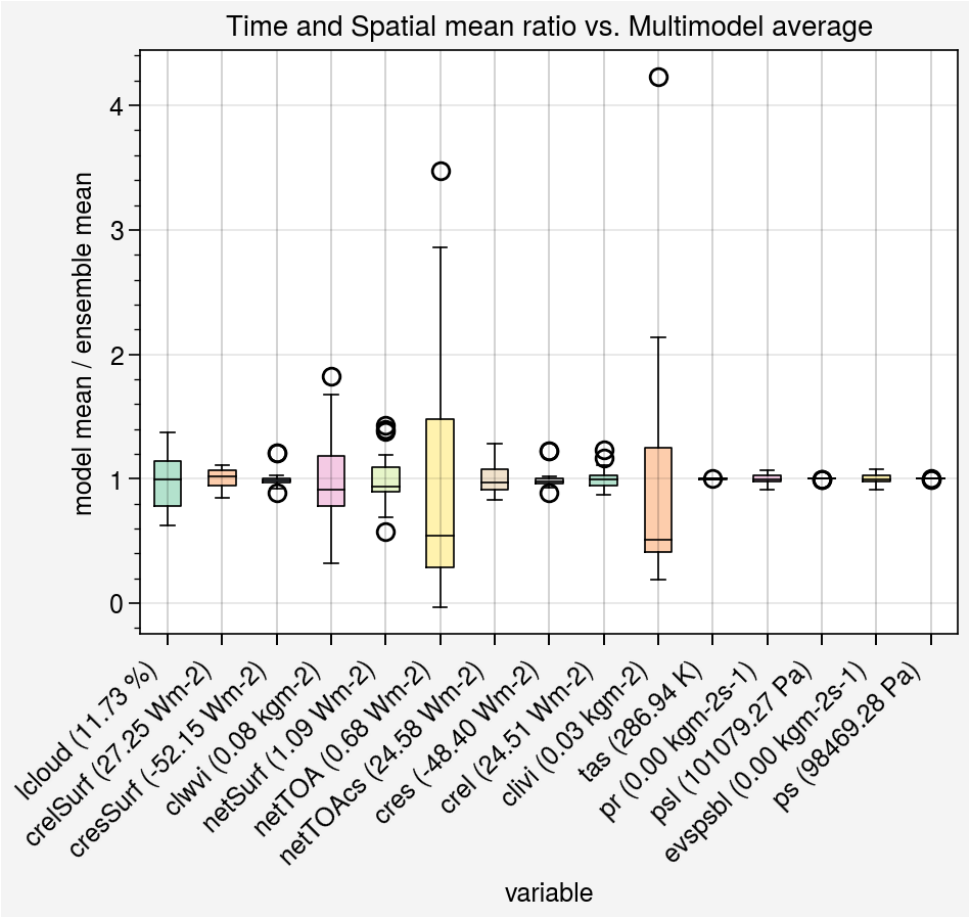

Figure 1: Box plot of spread of ESM global and time means of selected input, output, and constraint variables divided by their respective multi-model ensemble mean. Ensemble mean values shown in parentheses in the x-labels.

Each of those data hyper-cubes are preprocessed (repeat from Milestone 2 report) before ingestion into the hybrid model as follows:

1. **Remove seasonal cycle or "Deseasonalizing"**: We perform this process to remove any trends in the season to prepare a seasonal stationary time series data.

2. **Remove trend or "Detrend"**: We fit a third degree polynomial to remove any trend in data over time. This removes secular trends (for example, rising temperatures as atmospheric CO_2 increases) and allows the model to be trained on fluctuations due to internal variability, rather than the forced response.

3. **Normalized anomalies**: The anomaly at each grid point is calculated relative to a running mean that is computed over a centered 30-year window for that grid point and month. Anomalies are normalized by dividing by the standard deviation of the anomaly over the same 30-year window for that grid point and month.

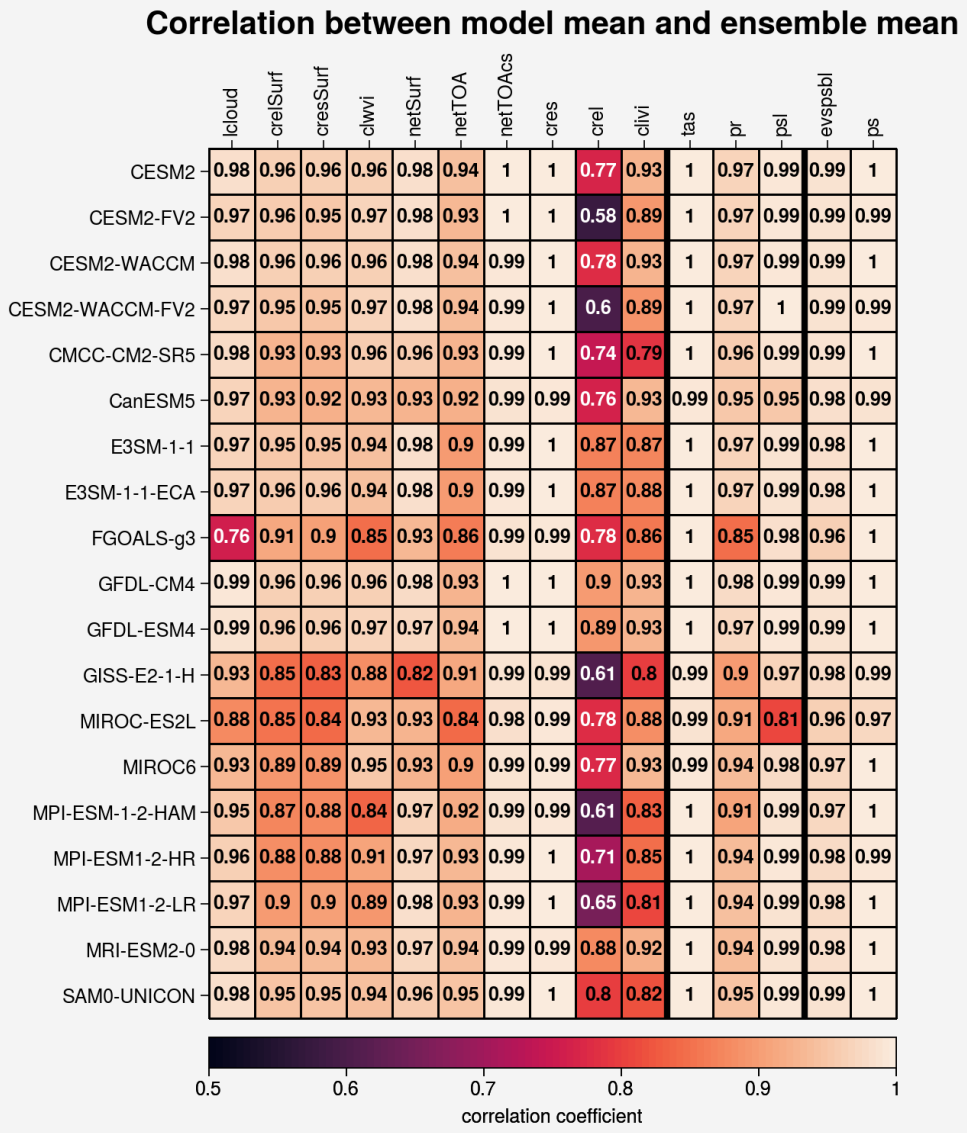

Figure 2: Pearson-R spatial correlations between ESM time average and ESM ensemble mean fields (for data remapped to level 5 Sphere-Icosahedral grid) across the models (rows) and variables (columns), showing the inter-ESM uncertainty in the climatologies of the selected input, output, and constraint variables.

To facilitate the training of the hybrid model that uses a *spherical grid* structure, these data hypercubes are transformed into icosahedral grids at levels 5 and 6, which lead to 10242 and 40962 data points, respectively (please refer to Milestone 2 report for details on spherical grids and levels). We also compressed the datasets by changing the data types (e.g. float64 to float32). The preprocessing and grid transformation codes can be found here.

Data deliverable

All the preprocessed datasets are currently available on AWS S3 cloud storage. The total size of the preprocessed and compressed datasets is around 150 GB. Datasets of specific models, or sub-selection of climate variables from our input and output hyper-cubes can be made available upon request.

Ongoing Work:
We continue to preprocess observational reanalysis datasets to assess the hybrid model performance and finetune the same to produce realistic simulations of the climate system response. In addition, we are developing scenarios with targeted cloud perturbations to generate tailored ESM runs. These datasets will be used to test and further fine-tune the model in Phase 2.

Model Performance Metrics

In Phase 1, we are assessing the hybrid model performance on two main metrics: accuracy and speed of inference, compared to the run time of a conventional Earth system model. In general, we aim to achieve at least 80% model accuracy and at least 10^4 time acceleration compared to conventional Earth system model performance.

- To assess model accuracy, we will report MSE (Mean Squared Error) of the model between model predictions of output variables and the 'ground truth' (values from Earth system model output).
- To assess regional accuracy of the model, we have divided the regions as shown in Figure 3, which consists of the tropics, midlatitudes (northern hemisphere and southern hemisphere), Arctic, and Antarctic zones. In each zone, we will report the error metrics of land and ocean areas separately using a land-sea mask attribute.

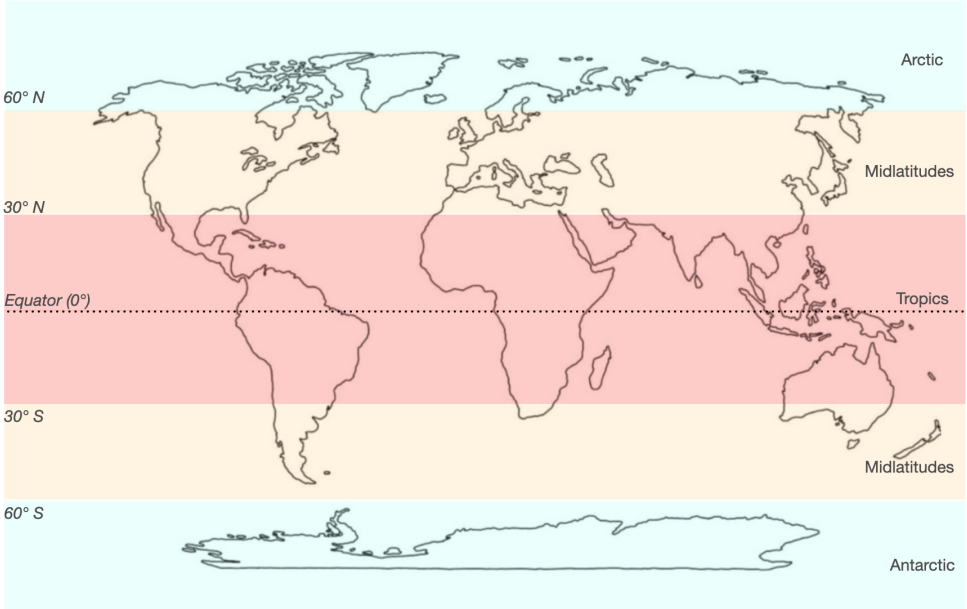

Figure 3: Regions for assessing region-wise metrics

- As we start including the physics constraints during model training, we will assess the impact on model performance for each additional physics constraint, as well as a combination of all the constraints. These will be reported for the entire model and the sub-regions.
- Finally, once the full model is trained and used for inference, we will record the time taken to obtain the output predictions for a given input variable. This will be compared with the time taken to run different Earth System models (CESM, E3SM, etc.)

Hybrid Model Development Update

Our previous milestone report explained different spherical sampling techniques to generate icosahedral or healpix grids. The datasets are interpolated onto these grids to train the spatial module: a Spherical U-Net

network (see more details here). We first trained the spatial module using CESM2-FV2 model ensemble output. Initial results of the model are shown in Figure 4.

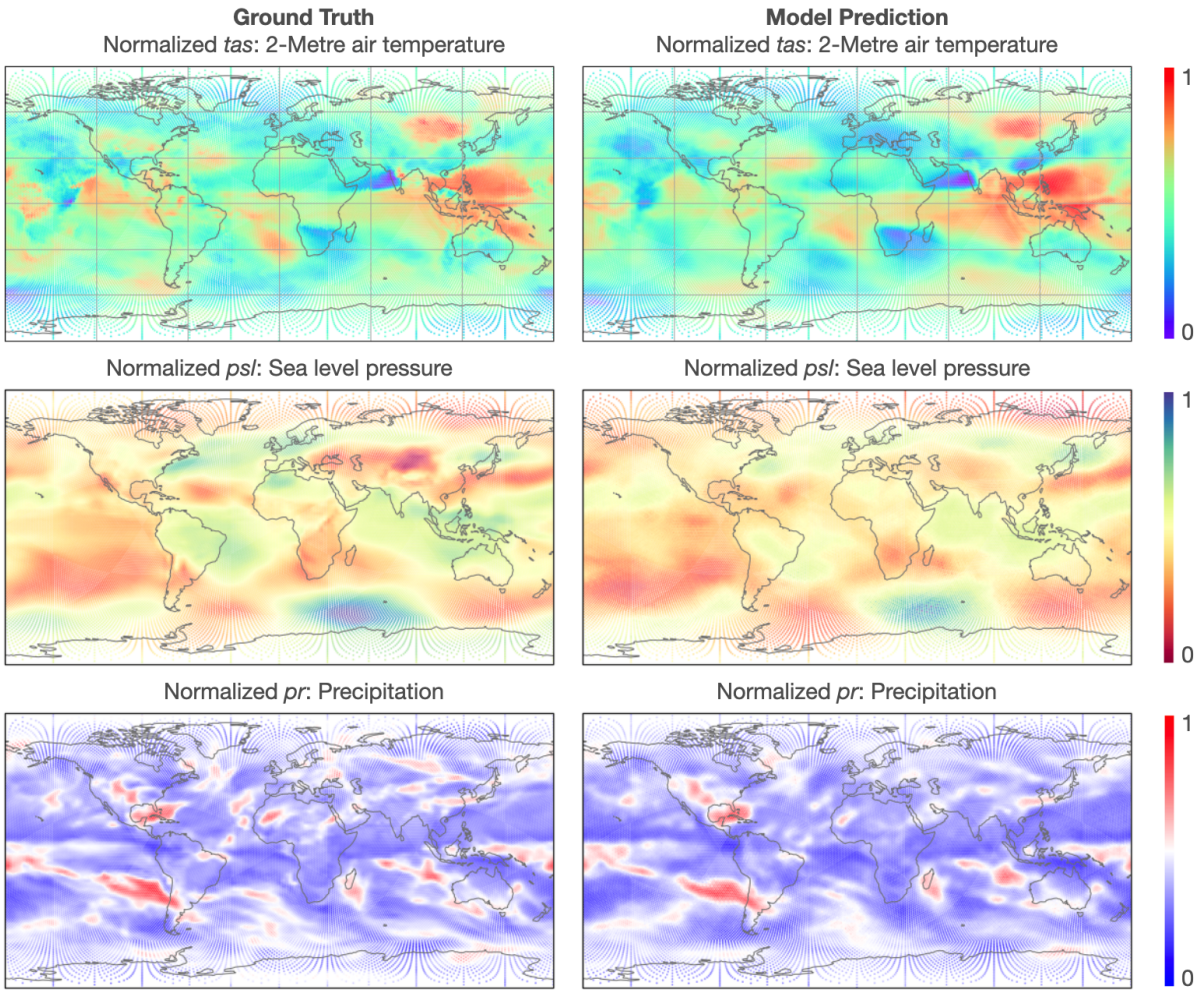

Figure 4: Initial Results: Ground truth (ESM model output) vs. AIBEDO hybrid model prediction

We observe that the hybrid model is generally good at capturing the patterns for all three output variables. However, the model output is poor in certain regions (e.g., tropics, due to high variability). To investigate this further, we plotted the errors split across regions (tropics, extratropics, arctic, antarctic, and over land and ocean) in Figure 5. We see that tropics have a larger interquartile range of errors, confirming high variability of the relationship between input and output variables. To mitigate this, we have identified physics constraints specific to tropical regions (explained in the Physics Constraints subsection), which will be incorporated in our next iterations.

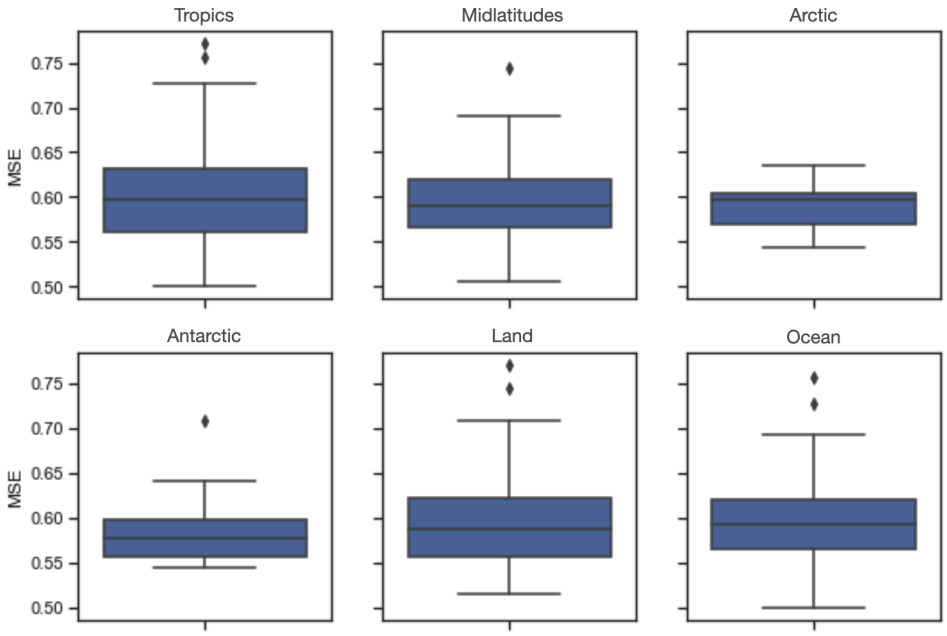

Figure 5: Region-wise Mean-Squared Error Boxplots for 2-Metre Air Temperature Prediction

While a stand-alone spatial model is good at capturing the spatial context between two sets of variables, this may not be adequate to capture climate responses which are typically time-dependent. For instance, top-of-atmosphere radiative anomalies drive climate responses that occur over a range of time scales. The fast response to radiative perturbations occurs over days and months following initial application of the forcing, while the slow response may occur over years.

In our use-case, we are predicting changes in atmospheric temperature, sea level pressure and precipitation, due to any changes in cloud properties. Our spatial data points are generated as an aggregated "snapshot" of each month. While this may capture the sub-monthly changes, this typically misses the lag responses of output variables due to the large-scale effects of the climate system. To ensure that our model captures these effects, we are exploring techniques to combine spatial and temporal modules. Our ongoing work includes development of model architecture that uses spatiotemporal inputs to predict the output. This includes a śequence' of spatial snapshots of input variable for a given *time length*, n from X_0, X_1,...X_n to predict the output variable at Y_n.

Physics constraints

Zero-shot generalization to unseen initial conditions is an important objective of the AIBEDO model, in Phase II. An effective method to enforce convergence that guarantees an out-of-distribution generalization is to use physically consistent constraints on model outputs. To that effect, we propose the following functional relationships that will be implemented as Lagrangian multipliers to the model loss.

- **Climate energy budget**: In this constraint, the energy is budgeted, between the heat storage and radiative fluxes at TOA, on longer climate-relevant timescales.

$$\sum_{t}^{>1yr} \sum_{lat=90S}^{90N} \sum_{lon=180W}^{180E} (\Delta R^{TOA}_{lat,lon} - \lambda_{ECS} T_{lat,lon} \Delta A_{lat,lon}) = 0 \tag{1}$$

  where ΔR^{TOA} is heat storage, λ_{ECS} feedback constant, T is surface temperature and A is the area of the cell.

- **Tropical atmospheric energy budget**: Unlike the climatic counterpart, this atmospheric budget balances the contributions from upward net radiative heat flux at the TOA and SFC to the heat convergence in the tropics, at the model prediction timescales.

$$\sum_{lat=30N}^{30N} \sum_{lon=180W}^{180E} (LP - R_{TOA} + R_{SFC} + SH + Q)_{lat,lon} \Delta A_{lat,lon} = 0 \tag{2}$$

where L is the latent heat of vaporization, SH is the sensible heat flux. Functionally, R^{TOA} and R^{SFC} can be calculated as the sum of the long wave and shortwave radiation at the top of atmosphere (TOA) and surface (SFC).

- **Global moisture budget**: This simple relationship balances the contributions from precipitation and evaporation, summed over all grid points.

$$\sum_{lat=90S}^{90N} \sum_{lon=180W}^{180E} (P - E)_{lat,lon} \Delta A_{lat,lon} = 0 \tag{3}$$

where P is the precipitation and E is the evaporation.

- **Non-negative precipitation**: A simple yet relevant constraint is to ensure negative precipitation (P) values are set to zero during model training. This will then ensure maximum penalty for the erring grid point when compared to the ground truth data.

$$P \geq 0, lat \in [90S, 90N], lon \in [180E, 180W] \tag{4}$$

- **Global atmospheric mass budget**: Using first principles, surface pressure can be used as a proxy for the atmospheric density, and therefore mass. This simple constraint ensures mass consistency summed over all the grid points for a particular timestep.

$$\sum_{lat=90S}^{90N} \sum_{lon=190W}^{180E} (P_s)_{lat,lon} \Delta A_{lat,lon} = 0 \tag{5}$$

where P_s is the surface pressure.

For example, a code snippet for enforcing the positive precipitation constraint is shown below.

```
def precip_pos(output):
    # This function sets any negative precipitation values to zero to
    # maximize penalty during backpropagation
    # shape of model output [time, resolution, fields ]
    # fields are [surf temp, surf pressure, precip]

    precip = np.array(output[:, :, 2])

    precip[precip<0] = 0

    output[:, :, 2] = precip

    return output
```

As all the different constraints are added and updated, the overall loss, which will be used for layer-wise back-propagation, is functionally defined as below, where *constraint_loss* is the sum of all the proposed constraints above and the *model_loss* is the RMSE loss of the original model prediction.

```
overall_loss = model_loss + lambda * constraint_loss
```

The loss function codes can be found here.

Ongoing Work:
Our ongoing work includes: (a) development of spatiotemporal hybrid model to incorporate lag response of climate effects, (b) including the coded physics constraints as part of the loss function in the hybrid model, and (c) postprocessing modules to interpret the results. We will then test the model performance against two datasets: earth system model outputs and observational reanalysis data. The former will determine if the model performs as intended for a given dataset, and the latter will determine which ESM (in the list shown in Table 1) is best captures observed relationships between clouds, the circulation, and surface climate with internal variability. Details of next steps are explained in the following section.

Next Steps

AIBEDO model training

We have split our AIBEDO model training in three stages covering phases 1 and 2, to ensure to capture natural variability of climate response as close as possible to the real-world, and to reduce uncertainty in model predictions. Our models are essentially trained on ESM model outputs, and as mentioned in the Datasets section, there are substantial differences in the climatologies and variability of some of the chosen input variables across models (see Figure 1). Depending on how they are represented in the ESM model, these may introduce biases in the hybrid model. To develop a model with cloud and atmospheric circulation response as close to the real-world observation, we are training the hybrid model in three stages as shown in Figure 6.

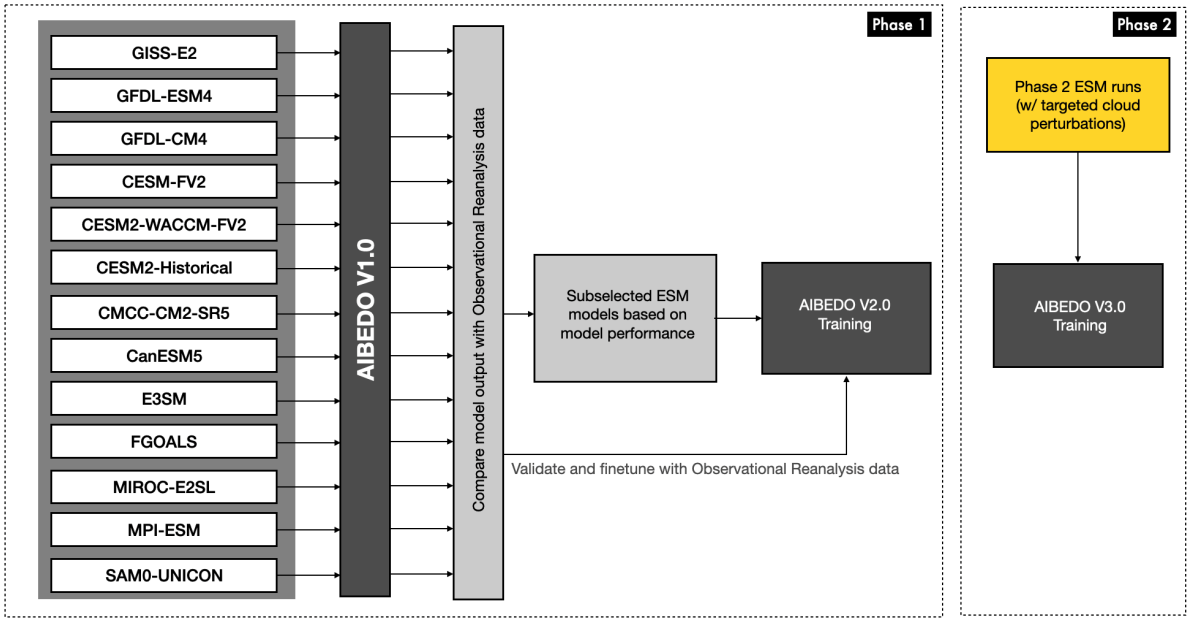

Figure 6: AIBEDO Model Training Stages

In stage 1, we will individually train the AIBEDO model for each ESM dataset and test its performance against observational reanalysis data. The error difference will be used to subselect the datasets that have the least error values. If needed, we will obtain more ensembles from these selected ESM models and preprocess them as training data in the next stage. In stage 2, we will train the AIBEDO model with the improved ESM model data. At this stage, we will be validating and finetuning the model with observational reanalysis data to account for any gaps evident in the initial tuning to capture the natural variability. This will be provided as a major deliverable towards the end of Phase 1. In Phase 2, we will utilize the ESM model output with targeted cloud perturbation experiments for refining our trained hybrid model in phase 1. AIBEDO V2.0 will be first validated against these datasets and will be finetuned to produce AIBEDO

V3.0. We will also reconsider including additional physics constraints to capture perturbation response adequately.

Visual Analysis (VA) system

The VA system will assist the end-users in performing different posthoc analyses and exploratory studies on the data and the pre-trained hybrid model. Our system will have the following outlined features:

- **Model prediction visualization:** the VA system will allow the users to visualize the input data sequences and the corresponding model predictions in juxtaposed panels. We will use a side view to convey the individual model confidence for the predicted results using a dropout-based bayesian approximation of the DL model.
- **What-if analysis:** Using the pre-trained hybrid model, we will extract the attribution maps for individual input fields (spatiotemporal) to understand the importance of input regions of different model outputs. We will compare three different attribution schemes: (a) saliency (sensitivity), (b) integrated gradient, and (c) layerwise-relevance propagation. Users can interactively select the region of interest in the output spatial domain (e.g., region of very low precipitation) to derive the corresponding attribution maps of the input fields, thus reflecting the possible input spatiotemporal features contributing to the selected output values. Our VA system will have interactive controls to update the values in selected input fields (e.g., cloud radiative forcings) to create new input fields for what-if analysis. We can compare the previous model predictions with the updated model predictions within the VA system to facilitate exploratory analysis with different unseen scenarios for domain experts. To enable interactive update of underlying data distribution of a user-selected input region, we will adopt three input update schemes:

  1. *Uniform update*: This naive scheme will let the users uniformly adjust (increase/decrease) the current data values of the user-selected spatial locations.
  2. *Attribute-weighted update*: Instead of uniformly updating the values at all selected locations, in this scheme, we will increase or decrease the original values based on the attribution scores of each location. This will help us distribute the effect of updating the input fields to salient regions of the field.
  3. *Radial update*: This scheme will use a radial kernel to distribute the updated values across the selected locations. The center of the selected regions will have the largest influence, while the effect will fade as we move away from the center.

  These different schemes will help the users perform what-f analysis studies at varying levels of detail.