Statistics of extreme events in climate models via coarse-scale simulations and machine learning correction operators based on nudged datasets

A. Charalampopoulos[1*], S. Zhang[2], B. Harrop[2], L. R. Leung[2], and T. P. Sapsis[1*]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology
Cambridge, MA 02139, USA
[2]Pacific Northwest National Laboratory, Richland, WA 99354, USA

Abstract

This work presents a systematic framework for improving the predictions of statistical quantities for turbulent systems, with a focus on correcting climate simulations obtained by coarse-scale models. Specifically, failure to incorporate all relevant scales in climate simulations leads to discrepancies in the energy spectrum as well as higher order statistics. While high resolution simulations or reanalysis data are available, at least for short periods, they cannot be directly used as training datasets to machine learn a correction for the coarse-scale climate model outputs, since chaotic divergence, inherent in the climate dynamics, makes datasets from different resolutions incompatible. To overcome this fundamental limitation we employ coarse-resolution model (here we employ Energy Exascale Earth System Model, E3SM) simulations nudged towards high quality climate realizations, here in the form of ERA5 reanalysis data. The nudging term is sufficiently small to not "pollute" the coarse-scale dynamics over short time scales, but also sufficiently large to keep the coarse-scale simulations "close" to the ERA5 trajectory over larger time scales. The result is a "compatible" pair of the ERA5 trajectory (used as output training data) and the weakly nudged coarse-resolution E3SM output that is used as input training data to machine learn a correction operator. We emphasize that the nudging step is used only for the training phase. Once training is complete, we perform free-running coarse-scale E3SM simulations without nudging and use those as input to the machine-learned correction operator to obtain high-quality (corrected) outputs. The model is applied to atmospheric climate data with the purpose of predicting global and local statistics of various quantities of a time-period of a decade. Using ERA5 datasets that are not employed for training, we demonstrate that the produced datasets from the ML-corrected coarse E3SM model have statistical properties that closely resemble the observations. In particular, the corrected coarse-scale E3SM output closely captures the non-Gaussian statistics of quantities such as temperature, wind speed and humidity, as well as the frequency of occurrence of extreme events, such as tropical cyclones and atmospheric rivers. We present thorough comparisons and discuss limitations of the approach.

*Corresponding authors: alexchar@mit.edu, sapsis@mit.edu
Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Problem Formulation

Accurate statistical climate predictions require high-fidelity simulations that come with large computational cost. As a result, improving upon the predictions of coarse-scale climate models has become a critical goal in order to develop credible climate scenarios. The developed method aims to augment the accuracy of statistical properties of coarse-scale free-running (i.e. without the influence of observations which are obviously not available for future projections) climate models, via corrections based on past data. Despite recent successes in using online correction terms in the evolution equations of the model (Yuval and O'Gorman 2020; Yuval, O'Gorman, and Hill 2021; Sanford et al. 2022; Charalampopoulos and Sapsis 2022a), many such implementations face severe stability issues. To circumvent this obstacle, a non-intrusive approach is developed. Hence, after a free-running coarsely resolved climate output has been generated, the hybrid approach corrects the model output in a post-processing manner. A reference dataset is selected for testing the effectiveness of the scheme, the ERA5 dataset (Hersbach et al. 2020). For the generation of coarse-scale climate data, the atmospheric component of the Energy Exascale Earth System Model (E3SM) is used. In particular, version 2 of the E3SM Atmospheric Model (EAMv2) (Dennis et al. 2012; Taylor, Cyr, and Fournier 2009; Golaz et al. 2022). Appropriate boundary conditions over the Earth's surface are prescribed (Oleson et al. 2013). Simulations are run on an unstructured grid of approximately $1^{\mathrm{o}}(\sim 110[\mathrm{km}])$ resolution per sigma-level and 72 levels along the vertical direction. The vertical levels extend from the Earth surface up to altitude of about $64[\mathrm{km}]$, corresponding to $\sim 0.1[\mathrm{hPa}]$. The evolution model equations for the coarse-scale model have the form

$$\frac{\partial \mathbf{X}_c}{\partial t} = \mathcal{D}\left(\mathbf{X}_c\right) + \mathcal{P}\left(\mathbf{X}_c\right), \tag{1}$$

where $\mathbf{X}_c = (U, V, T, Q)$ represent the set of coarsely-resolved system variables, $\mathcal{D}$ is the operator containing the dynamics of the system (Zhang and McFarlane 1995; Golaz, Larson, and Cotton 2002) and $\mathcal{P}$ is the operator concerning the physics of the system (Morrison and Gettelman 2008; Liu et al. 2016; Mlawer et al. 1997). Variables (U, V) correspond the zonal and meridional components of wind velocity, T is wind temperature and Q is specific humidity. From

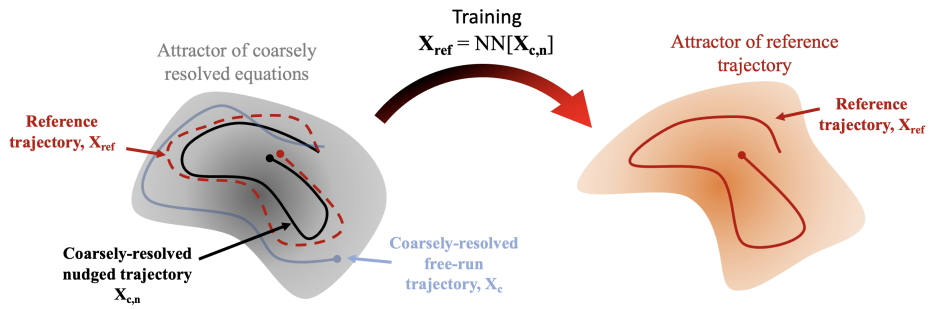


Figure 1: **Description of the method** that learns a map between the attractor of the coarsely-resolved equations and the attractor of the reference trajectory. Left: the red dashed curve represents a reference trajectory (here ERA5). The black curve is a coarsely-resolved nudged trajectory towards the reference trajectory. The blue curve is the free-run coarsely-resolved trajectory that is not used for training (shown for reference). Right: the target attractor and the target trajectory (red), same as the dashed curve shown at the left plot. For training we use the coarsely-resolved nudged trajectory as input and the reference trajectory as output to machine learn a map between them. After we obtain the map we use as input coarsely-resolved free-run simulations (blue) and obtain a trajectory that accurately captures the shape of the target attractor.

here on now, the coarse-scale free-running dataset will be labeled as CLIM and will be denoted as $\mathbf{X}_c$.

For reference data, denoted as $\mathbf{X}^{\text{ref}}$, ERA5 reanalysis data is used, which is projected onto the coarse unstructured grid of EAMv2. The datasets discussed herein contain information from 2007-2017. For this timescale the studied climate systems can be assumed to be in a statistical steady state.

Since the goal of the approach is to correct the long-time statistics of coarse-scale climate simulations in a post-processing manner, it is important to isolate the main discrepancies between the coarse-scale simulations and the reference data that are responsible for these differences. In general, discrepancies between two turbulent simulations, one high-fidelity (i.e. reference) and one free-running coarse, can be grouped into two categories: (i) discrepancies due to chaotic divergence; (ii) discrepancies due to deformation of the attractor due to coarse-scale resolution.

Chaotic divergence is an intrinsic property of turbulent systems. It can be observed even between two solutions of the same dynamical system, with ever slightly different initial conditions. It is a manifestation of the fact that by definition, at least one of the eigenvalues of the linear part of the system is positive. As a result, infinitesimal energy transferred to perturbations along these directions will result in finite magnitude perturbations. The system is allowed to equilibrate with the intervention of nonlinear terms that will transfer this energy from the unstable perturbations to stable ones. However, the two deviating trajectories will still remain on the same attractor and thus retain the same statistical properties. Therefore these chaos-induced discrepancies should not contribute to the correction scheme for long-time statistics.

On the other hand, difference in long-time statistics implies a different statistical steady-state and thus different attractors. These intrinsic dynamical differences between the simulations produce energy discrepancies in various scales between the produced datasets. It is exactly these corrections we aim to learn and fix.

Given the two previous observations, it is clear that it is not possible to use a dataset of free-running climate simulation (CLIM) and one of ERA5 and try to machine learn a map between the two, i.e. a map that takes as input a CLIM timeseries and produces as output an ERA5 timeseries. To eliminate the problematic component, i.e. chaos-induced divergence we design a new CLIM dataset (we call it nudged CLIM and denote it as $\mathbf{X}_{c,n}$). Ideally, one can produce a dataset that is preserving the coarse-scale behavior of the climate model but does not suffer from the chaos-induced divergence with the ERA5. To this end, the concept of nudging, that has been used extensively in the context of data assimilation (Sun et al. 2019; Zhang et al. 2022), is employed. Specifically, we utilize the EAMv2 solver (generator of the CLIM dataset) with an extra term, the nudging term, that is 'pulling' the CLIM solution close to the ERA5 solution:

$$\frac{\partial \mathbf{X}_{c,n}}{\partial t} = \mathcal{D}\left(\mathbf{X}_{c,n}\right) + \mathcal{P}\left(\mathbf{X}_{c,n}\right) + \mathcal{N}\left(\mathbf{X}_{c,n}; \mathbf{X}^{\text{ref}}\right), \quad (2)$$

where the relaxation term $\mathcal{N}$ is called the nudging tendency and it corrects the coarse-scale solution based on the ERA5 reference solution. In this study, the nudging tendency $\mathcal{N}$ is given by the algebraic term

$$\mathcal{N}\left(\mathbf{X}_{c,n} - \mathbf{X}^{\text{ref}}\right) = -\frac{1}{\tau}\left(\mathbf{X}_{c,n} - \mathcal{H}\left[\mathbf{X}^{\text{ref}}\right]\right). \quad (3)$$

Parameter τ is a relaxation timescale that has a large value (so that $1/\tau$ is small compared with the other terms in the equation), and $\mathcal{H}$ is an operator that maps $\mathbf{X}^{\text{ref}}$ to the coarse resolution. A schematic of the proposed mapping learned during training can be seen in fig. 1.

The resulted nudged trajectory (black curve on the right panel) is subjected to this very small perturbation, the nudging term that is keeping it close to the reference trajectory, i.e. the ERA5 trajectory (red dashed line). Moreover, because the overall magnitude of the nudging term is very small the long-time statistics of the coarsely resolved before-nudge trajectory should be close to that of the free-running coarsely-resolved trajectory that starts from the

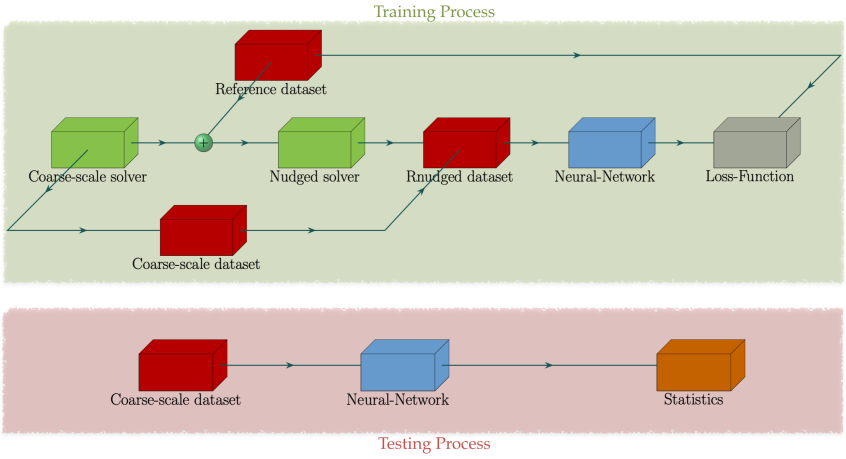


Figure 2: Schematics of the training process (top) and testing process (bottom), for the non-intrusive hybrid method.

before-nudge state (shown with blue color). The latter will naturally diverge from the ERA5 if not continuously being nudged due to chaotic properties, even if it was initiated very close to ERA5. Having the before-nudge trajectory we can now use it to machine learn a scheme that will map it to the reference trajectory, i.e.the ERA5. That is the basic approach of our framework. We emphasize that nudging is used ONLY for the generation of training data. Once the map has been trained we will feed it with free-run CLIM simulations (i.e. free-run coarse-scale climate simulations without nudging) to obtain outputs that have corrected long-time statistics, i.e. represent the target attractor accurately.

The resulting training and testing process are described in fig. 2. During training, the Nudged EAMv2 solver is used to produce training data. After a spectral correction of the data (described in the next section), the resulting R-nudged dataset is used as input for the neural network. The neural network then learns a mapping between the reference ERA5 data and the input R-nudged dataset. The neural network used is described in detail in subsection 'Neural Network Architecture'. During testing, EAMv2 is used to generate a free-running coarse-scale dataset. This dataset is used as input to the trained neural network which produces a corrected dataset with the desired statistics. *Hence, during testing, the model is not assessed in its ability to mimic the reference data snapshot-by-snapshot but by its ability to learn its underlying statistics*.

Revisiting the nudging procedure, parameter τ is chosen so the nudged solution $\mathbf{X}_c$ satisfies two properties: (a) it reduces the divergence of the nudged simulation from the reference solution (i.e., ERA5) $\mathcal{H}\left[\mathbf{X}^{\text{ref}}\right]$, i.e. allowing for a generalizable mapping between the two datasets; (b) it resembles the statistical properties of the coarse-scale free-running simulation. The second property is important in the context of machine learning to ensure that the learned mapping during training will be applicable while testing using CLIM data. This implies that the attractor of the Nudged simulation has the same shape as the attractor of CLIM. However, no parameter τ can be found that explicitly satisfies this condition. This is due to the arbitrariness of the algebraic form of the nudging term. While an algebraic term is easily implemented it yields a constant dissipation rate across all wavenumbers which are in general not consistent with the dynamics of the system. This leads to suppression of extreme events and thus statistics with less heavy tails. A remedy for this issue is shown in subsection 'Spectral Correction of Nudged Dataset', where the energy spectrum of nudged simulations is brought closer to that of CLIM. One other possible remedy is changing the form of the nudging term via an appropriate energy balance argument.

Spectral Correction of Nudged Dataset

As described in the previous subsection, the nudged dataset $\mathbf{X}_{c,n}$ is used during training to eliminate discrepancies due to chaotic divergence between input data and ERA5 reference data. However, for the learned mapping to be completely transferable to the free-running $\mathbf{X}_c$ testing data, the two datasets must have attractors that are as similar as possible. If not, these discrepancies will manifest themselves to the target of this work, i.e. tails of statistics of various quantities. This is a result of discrepancies in the energy spectrum of the nudged solution with respect to the coarse-scale solution. These energy spectra differences lead to different statistical steady-state behaviours of testing data $\mathbf{X}_c$ and training data $\mathbf{X}_{c,n}$, which inhibits the generalizability of a model (Shalev-Shwartz and Ben-David 2014).

To remedy the energy spectra differences between the testing input data $\mathbf{X}_c$ and training input data $\mathbf{X}_{c,n}$, a non-intrusive correction is proposed. The process is called 'Reverse Spectral Nudging' with its purpose being to match the energy spectrum of the nudged solution to that of the coarse-scale solution. This modification implies that the statistics of the steady state of the two input datasets are very close to one another and thus generalizability improves. Hence, while traditional nudging schemes correct the coarse-scale solution with data from the reference solution, the proposed scheme further processes the nudged data by matching its energy spectrum to that of the corresponding free running coarse-scale flow. The corrected nudged data is termed as

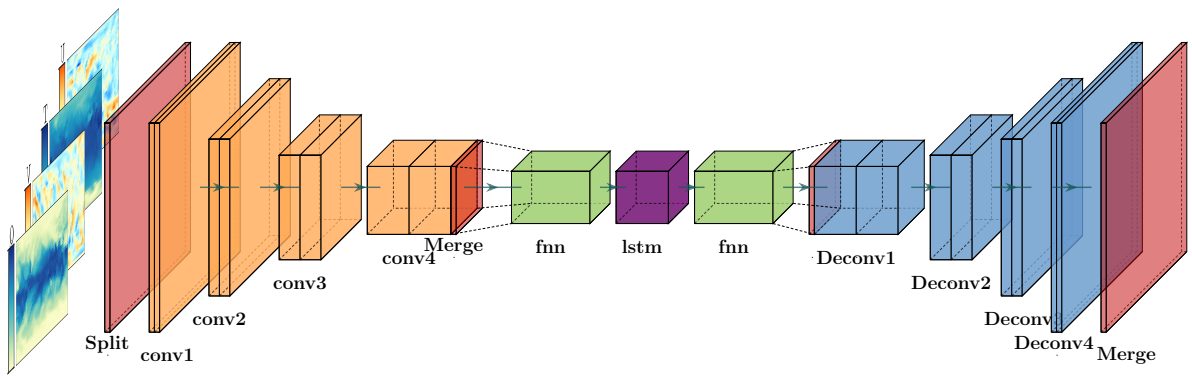


Figure 3: Neural network architecture of the non-instrusive model for a training on a particular sigma-level.

$\mathbf{X}_{c,rsn}$ and defined as

$$\mathbf{X}_{c,rsn}\left(\phi,\theta,t;k\right)=\sum_{m,n}R_{m,n}\{\hat{\mathbf{X}}_c(t)\}_{m,n}e^{i(m\phi+n\theta)},, \tag{4}$$

where $\{\hat{\mathbf{X}}_c(t)\}_{m,n}$ are the spatial Fourier coefficients of $\mathbf{X}_{c,n}$, ϕ and θ corresponds to longitute and latitude respectively, while parameter k denotes the number of the sigma-level. We also have

$$R_{k,l}=\sqrt{\frac{\mathcal{E}_{k,l}^{\text{coarse}}}{\mathcal{E}_{k,l}^{\text{nudge}}}}, \tag{5}$$

and

$$\mathcal{E}_{k,l}=\frac{1}{T}\int_0^T\hat{E}_{k,l}(t)\mathrm{d}t=\frac{1}{T}\int_0^T|\{\hat{\mathbf{X}}_c(t)\}_{k,l}|^2\mathrm{d}t, \tag{6}$$

where T is the duration of the available data.

Neural Network Architecture

In the current implementation, training is done on an level-by-level basis, where level here denotes atmospheric sigma-levels (Taylor et al. 2020). A schematic of the configuration for training on a particular layer is shown in fig. 3. The model receives as input the predictive variables $\mathbf{X}=\mathbf{X}(\phi,\theta,t;k)$, where ϕ is the longitudinal angle and θ the latitudinal one. Snapshots of the entire horizontal discretization of the layer are used. Afterwards, a custom "split" layer separates the input into non-overlapping subregions. These subregions are periodically padded via a custom padding process, tasked with respecting the spherical periodicity of the domain. Then, each subregion is independently passed through a series of convolutional layers. The purpose of this process is to extract anisotropic local features in each subregion.

Afterwards, the local information extracted from each subregion is concatenated in a single vector via a custom 'merge' layer. The global information is now passed through a linear fully-connected layer, that acts as a basis projection of the spatial data onto a reduced-order latent space. The latent space data are then corrected by a long short-term memory (LSTM) layer (Hochreiter and Schmidhuber 1997). Subsequently they are projected back to physical space via another linear fully-connected layer. Afterwards, global information is split into the same subregions of the input, and distributed to a series of independent deconvolution layers that upscale the data to the original resolution. Finally, a custom 'merge' layer gathers the information from each subregion and produces the final corrected snapshot.

The motivation behind using LSTM neural networks lies in their ability to incorporate (non-Markovian) memory effects into the reduced-order model. This ability stems from Takens embedding theorem (Takens 1981). This theorem states that given delayed embeddings of a limited number of state variables, one can still obtain the attractor of the full system for the observed variables. This approach is known to be capable of improving predictions of reduced-order models (Vlachas et al. 2018; Charalampopoulos and Sapsis 2022b; Wan et al. 2021). In addition to temporal nonlocality, the model is nonlocal in space. Note, that in terms of the LSTM layer, this information comes in the form of the latent space coefficients, which in general correspond to global modes that correspond to rows of the fully connected layer's matrix. Under the assumption that both fully-connected layers have linear activation functions, the model can be mathematically depicted as a basis projection. Hence, the fully connected layers act as projection schemes to (a) compress input data to a latent space of low dimensionality, and (b) project the LSTM prediction to physical space. The used loss function is a standard mean-square error (MSE) loss

$$\mathcal{L}=\alpha\sum_t\sum_\phi\sum_\theta\cos\left(2\pi\frac{\theta}{360}\right)\|\hat{\mathbf{X}}-\mathbf{X}^{\text{ref}}\|^2, \tag{7}$$

where α is a normalization coefficient. Each term in the sum is multiplied by a cosine that is a function of the latitude to showcase that the integration takes place over a sphere.

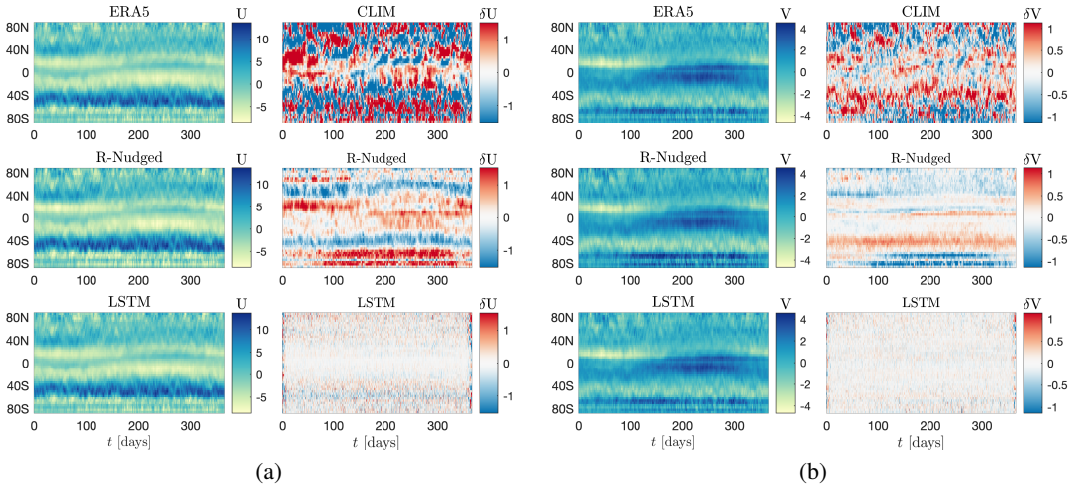


Figure 4: (a) Zonally averaged predictions for U (left column) and biases with respect to ERA5 predictions (right column). (b) Zonally averaged predictions for V (left column) and biases with respect to ERA5 predictions (right column). Results are shown for near-surface data.

If that term is absent, the model would over-emphasize on learning the corrections at the poles.

Application on E3SM Data

The numerical exploration of the proposed method begins with some validation results. Specifically, nudged data not used for training are used as input, with the purpose of checking whether the produced output has good accuracy. Training takes places over 1000 epochs and eq. (7) is used as the loss function. Training took place over the time-period 2007-2011 using only Nudged data as input, not CLIM data. Year 2012 was used for validation during training. In fig. 4, results regarding zonally averaged predictions of zonal and meridional velocities U and V respectively, are shown. The left column of subfigure (a) displays the zonally averaged predictions of zonal velocity U, for ERA5 reanalysis data, Nudged data and neural network predictions. The right column displays the biases compared to ERA5 predictions for CLIM, Nudged and the neural network predictions. Similarly, the left column of subfigure (b) displays the zonally averaged predictions for meridional velocity V, with the same biases displayed. Results are shown for sigma-level 71, i.e. the one closest to the surface of the earth. This level was then one CLIM predictions differed the most with respect to ERA5 reanalysis data. The results show that the neural network clearly is able to learn a correction for both velocity components.

Now, the model is tested on unseen free-running and coarse-scale climate models. This dataset (CLIM) is not nudged and thus does not include any ERA5 information. Since the free-running dataset diverges from ERA5 in terms of phases, specific extreme events cannot be studied, as they are absent from CLIM. Hence, long-time statistics are studied now.

In fig. 5, the predicted probability density functions (pdf) are shown for the four different predictive variables, (U, V, T, Q). Solid black lines correspond to ERA5 data, dashed black lines correspond to CLIM and green lines correspond to Nudged data. Red lines correspond to neural network predictions using Nudged data as input (i.e. training data). Blue lines correspond to neural network predictions using CLIM data as input (i.e. testing data). Results are shown for sigma-level closest to the earth's surface. Data are averaged over the time-period 2007-2017. The predictions of the neural network that was trained with Nudged data, significantly improve the prediction of the tails. This can be seen both for the pdfs of the well-predicted quantities (U, V) by CLIM as well as for (T, Q), two quantities whose tails are not well predicted by CLIM. Furthermore, the difference in the predicted pdf when using the training data and free-running data is insignificant, showcasing the ability of the model to generalize beyond training data. This is a result of the pdf of the implemented spectral corrections to the Nudged data, making them display very similar statistics to that of the free running coarse data. This property allows for smoother transfer learning between data sets.

We now move to predict statistics for a derived integral quantity, in particular, mean integrated vapor transport (IVT) over the period 2007-2017. Since IVT is strongly anisotropic, extracting local features, and especially atmospheric rivers, is vital. Therefore the local convolutions derived from each subregion are important for its correction estimation. 25 subregions are used in this numerical test, using different convolutions on each one (i.e. extract local features). In fig. 6, the first row corresponds to ERA5 predictions. The other rows correspond to biases with respect to ERA5. As expected, the neural network exhibits very small biases when Nudged data are used as input. Results for the nudged simulation are not shown for page limit reasons. However, the model significantly decreases biases

even when CLIM data are used as input. In addition, it is able to decrease the root mean-squared error (RMSE) below that of the Nudged dataset (which was measured at 0.088), a simulation that exploits ERA5 information at every time-step.

The final numerical result involves the statistics of tropical cyclones for the time-period 2007-2017. For tracking tropical cyclones, the software package TempestExtreme is used (Ullrich and Zarzycki 2017; Ullrich et al. 2021). The following steps are used to track cyclones: (i) Find local minima of sea-level pressure (SLP). (ii) Eliminate smaller minima within 2 great-circle distance (gcd) degrees. (iii) Check that SLP raises by 200 [Pa] within 8 gcd. (iv) Check that temperature at 400 [mbar] drops by 0.4 [K] within 8 gcd. (v) Check that velocity is higher than 10 [m/sec] for 8 snapshots. (vi) Check that geopotential height is larger than 100 for at least 8 snapshots. (vii) Check that phenomenon lasts at least 54 hours, with a max gap of 24 hours.

During training, the neural network will track dissipated cyclonic structures in the Nudged simulations, that correspond to tropical cyclones in ERA5. It will then amplify them allowing them to be recognized as tropical cyclones. As a result, if the cyclonic structures present in CLIM are more dissipated than in Nudged, the learned mapping will not be transferable. To overcome this issue, we post process the SLP values of Nudged data. To that end the following conditional means are computed

$$c(\phi,\theta) = \frac{1}{\mathrm{TC}(\phi,\theta)} \sum_{t=1}^{N_t} \delta SLP(t,\phi,\theta), \tag{8}$$

where TC is the tropical cyclone density over the globe. Then, the following spatially dependent coefficient is computed

$$R(\phi,\theta) = \frac{c^{\mathrm{c}}(\phi,\theta)}{c^{\mathrm{c,n}}(\phi,\theta)}. \tag{9}$$

Finally, the SLP Nudged data are corrected as follows

$$SLP^{\mathrm{c,rnsn}} = \begin{cases} R(\phi,\theta)SLP^{\mathrm{c,n}} + P_0, & SLP < P_0 \\ SLP^{\mathrm{c,n}}, & SLP \geq P_0 \end{cases}. \tag{10}$$

Results are shown in fig. 7. In terms of total number of tropical cyclones predicted, the neural network corrections are much better than the ones predicted by the Nudged dataset and CLIM. Nudged data predict 317 tropical cyclones, while CLIM predicts 305. On the other hand, the neural network corrections predict 411 tropical cyclones when using Nudged data and 404 tropical cyclones when using CLIM data. These predictions are much closer to the 488 cyclones predicted from ERA5 data. By focusing on the predictions when using CLIM as input, we notice that barely any new cyclones are predicted in the Atlantic. Looking further into this issue it is our belief that this is a problem of CLIM, where it does not generate enough vorticity over the Atlantic for tropical cyclones to form.

Limitations

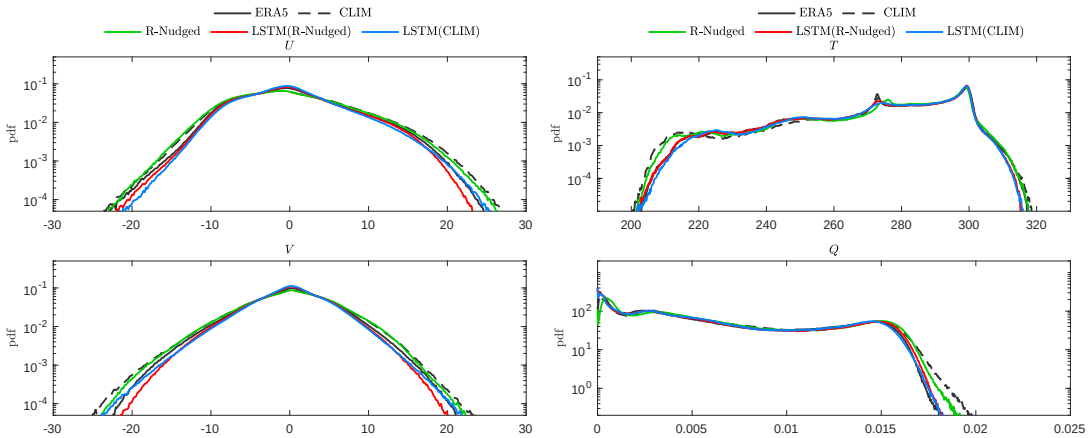
While the proposed methodology was demonstrated to be effective for the prediction of a multitude of climate metrics, some limitations of the current setup should be stated. First, the approach works well under the assumption that the climate is in a statically steady steady, for which a mapping can be learned through the proposed training scheme. Hence, testing the model in simulations where the climate undergoes a transitory phase may hinder its performance, unless similar time intervals are included during training. This is particularly true if the transition is not captured at all by the coarse-scale model. Furthermore, the requirement for reference data (in this case ERA5 reanalysis data), makes the effectiveness of the model unknown under future climate scenarios with drastically different forcings. For such runs to be included in training, high-fidelity simulations would have to be used as reference and nudged towards them. This limitation however is true for online data-driven correction schemes as well since most such models lack concrete error bounds for out-of-sample predictions. Finally, for the application of the scheme to dynamical systems broadly, there is no guarantee that a nudged simulation exists that follows the reference data closely while maintaining the statistics of the coarse simulation. In fact, this is not the case for climate models. However, small discrepancies can be amended via the proposed extra step of the spectral corrections. It should be noted yet, that there is no guarantee the process will work in a case where the deviations from these assumptions are significant.

Conclusions

We have formulated and assessed a data-informed hybrid scheme for accurately computing the statistics of climate models. The method employs a nudged solver during training with appropriate spectral corrections to the produced dataset. During testing, the model is assessed on free-running coarse-scale climate models. The approach was applied to realistic atmospheric climate data. Free-running EAMv2 simulations were used as baseline while ERA5 reanalysis data were employed as reference truth. First, the ability of the model to predict global statistics of horizontal velocities (U, V), temperature T and specific humidity Q was assessed. In all cases, the data-informed approach produced results in good agreement with reference ERA5 predictions. Furthermore, the model was tasked with predicting mean IVT over the period 2007-2017, a highly anisotropic quantity. Results are again in good agreement with ERA5, with the corrected dataset having a smaller mean-square error than a nudged dataset. Finally, the model was used to predict statistics of tropical cyclones throughout the globe. The model's ability to correct the number of predicted cyclones over the Pacific was demonstrated.

Acknowledgments

This research has been supported by the DARPA grant HR00112290029 with the program 'AI-assisted Climate Tipping-point Modeling' under the program manger Dr. Joshua Elliott. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract. We also thank the two anonymous referees for their constructive comments and feedback.



(a) (b)

Figure 5: Predicted pdfs for (a) horizontal velocity components U and V; (b) temperature T and specific humidity Q. Solid black lines correspond to ERA5 data, dashed black lines correspond to CLIM and green lines correspond to Nudged data. Red lines correspond to neural network predictions using Nudged data as input (i.e. training data). Blue lines correspond to neural network predictions using CLIM data as input (i.e. testing data). Results are shown for near-surface data.

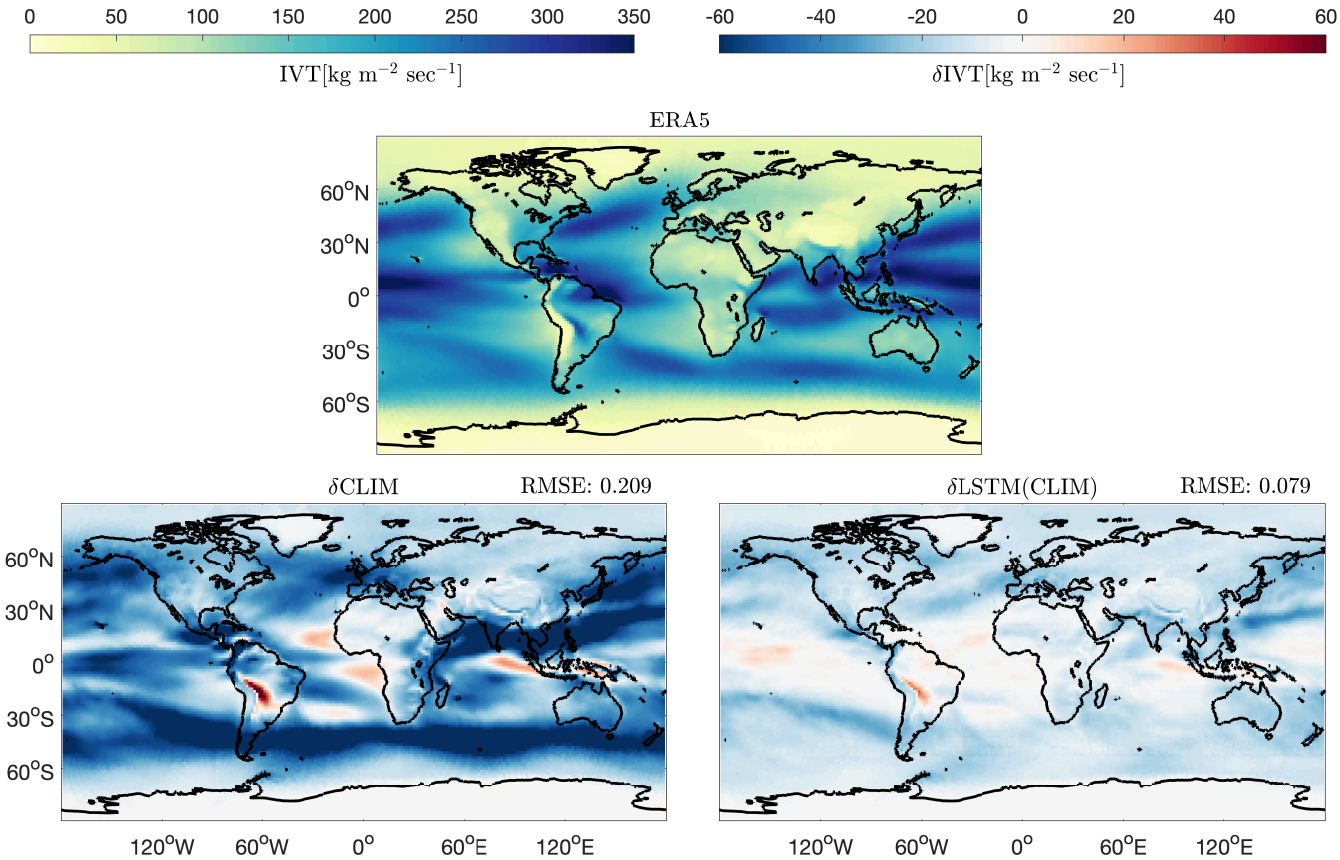


Figure 6: IVT predictions averaged over the period 2007-2017. Mean IVT predictions are shown in the top row for ERA5 data. Biases from ERA5 predictions are shown for, the CLIM free-running dataset, together with corrected results via our non-intrusive approach. A 25-subregion partition was used for these results.

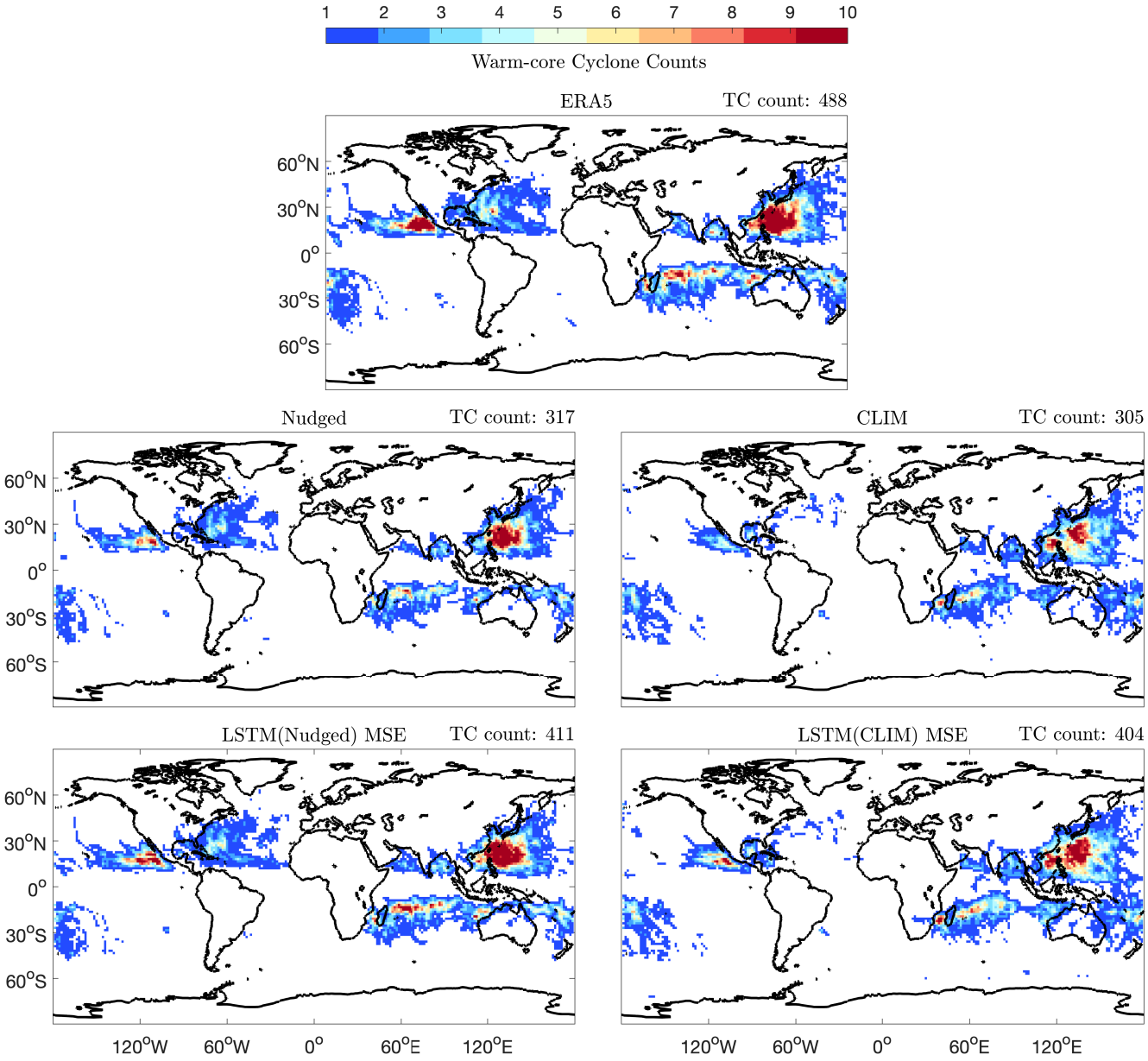


Figure 7: Tropical cyclone counts over the period 2007-2017. Results are derived using ERA5 datasets, CLIM free-running datasets and CLIM datasets corrected via our non-intrusive LSTM approach. Cyclones are tracked via the TempestExtremes software.

References

Charalampopoulos, A.; and Sapsis, T. 2022a. Uncertainty quantification of turbulent systems via physically consistent and data-informed reduced-order models. *Physics of Fluids*, 34(7): 075120.

Charalampopoulos, A.-T. G.; and Sapsis, T. P. 2022b. Machine-learning energy-preserving nonlocal closures for turbulent fluid flows and inertial tracers. *Physical Review Fluids*, 7(2): 024305.

Dennis, J. M.; Edwards, J.; Evans, K. J.; Guba, O.; Lauritzen, P. H.; Mirin, A. A.; St-Cyr, A.; Taylor, M. A.; and Worley, P. H. 2012. CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *The International Journal of High Performance Computing Applications*, 26(1): 74–89.

Golaz, J.-C.; Larson, V. E.; and Cotton, W. R. 2002. A PDF-based model for boundary layer clouds. Part I: Method and model description. *Journal of the atmospheric sciences*, 59(24): 3540–3551.

Golaz, J.-C.; Van Roekel, L. P.; Zheng, X.; Roberts, A. F.; Wolfe, J. D.; Lin, W.; Bradley, A. M.; Tang, Q.; Maltrud, M. E.; Forsyth, R. M.; et al. 2022. The DOE E3SM Model Version 2: overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, 14(12).

Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730): 1999–2049.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Liu, X.; Ma, P.-L.; Wang, H.; Tilmes, S.; Singh, B.; Easter, R.; Ghan, S.; and Rasch, P. 2016. Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model. *Geoscientific Model Development*, 9(2): 505–522.

Mlawer, E. J.; Taubman, S. J.; Brown, P. D.; Iacono, M. J.; and Clough, S. A. 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102(D14): 16663–16682.

Morrison, H.; and Gettelman, A. 2008. A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests. *Journal of Climate*, 21(15): 3642–3659.

Oleson, K.; Lawrence, D.; Bonan, G.; Drewniack, B.; Huang, M.; Koven, C.; Levis, S.; Li, F.; Riley, W.; Subin, Z.; et al. 2013. Technical description of version 4.5 of the Community Land Model (CLM)(Technical Note No. NCAR/TN-503+ STR). Boulder, CO: National Center for Atmospheric Research Earth System Laboratory.

Sanford, C.; Kwa, A.; Watt-Meyer, O.; Clark, S.; Brenowitz, N.; McGibbon, J.; and Bretherton, C. 2022. Improving the predictions of ML-corrected climate models with novelty detection. *arXiv preprint arXiv:2211.13354*.

Shalev-Shwartz, S.; and Ben-David, S. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Sun, J.; Zhang, K.; Wan, H.; Ma, P.-L.; Tang, Q.; and Zhang, S. 2019. Impact of nudging strategy on the climate representativeness and hindcast skill of constrained EAMv1 simulations. *Journal of Advances in Modeling Earth Systems*, 11(12): 3911–3933.

Takens, F. 1981. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, 366–381. Springer.

Taylor, M. A.; Cyr, A. S.; and Fournier, A. 2009. A non-oscillatory advection operator for the compatible spectral element method. In *International Conference on Computational Science*, 273–282. Springer.

Taylor, M. A.; Guba, O.; Steyer, A.; Ullrich, P. A.; Hall, D. M.; and Eldred, C. 2020. An energy consistent discretization of the nonhydrostatic equations in primitive variables. *Journal of Advances in Modeling Earth Systems*, 12(1): e2019MS001783.

Ullrich, P. A.; and Zarzycki, C. M. 2017. TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, 10(3): 1069–1090.

Ullrich, P. A.; Zarzycki, C. M.; McClenny, E. E.; Pinheiro, M. C.; Stansfield, A. M.; and Reed, K. A. 2021. TempestExtremes v2. 1: a community framework for feature detection, tracking, and analysis in large datasets. *Geoscientific Model Development*, 14(8): 5023–5048.

Vlachas, P. R.; Byeon, W.; Wan, Z. Y.; Sapsis, T. P.; and Koumoutsakos, P. 2018. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2213): 20170844.

Wan, Z. Y.; Dodov, B.; Lessig, C.; Dijkstra, H.; and Sapsis, T. P. 2021. A data-driven framework for the stochastic reconstruction of small-scale features with application to climate data sets. *Journal of Computational Physics*, 110484.

Yuval, J.; O'Gorman, P. A.; and Hill, C. N. 2021. Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters*, 48(6): e2020GL091363.

Yuval, J.; and O'Gorman, P. A. 2020. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, 11(1): 3295.

Zhang, G. J.; and McFarlane, N. A. 1995. Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere-ocean*, 33(3): 407–446.

Zhang, S.; Zhang, K.; Wan, H.; and Sun, J. 2022. Further improvement and evaluation of nudging in the E3SM Atmosphere Model version 1 (EAMv1). *Geoscientific Model Development Discussions*, 1–37.