Enhancing Weather Predictions with Hybrid Data and Physics-Based Models
(CSE 8803 SML 23 Fall, Group: CliML)

Abstract

At present the state-of-the-art in global weather modeling resides in two camps: numerical weather prediction (NWP), based on scientific models, and data-driven weather prediction, based on parametric models. The purely data-driven model FourCastNet offers several advantages over NWP models, including much faster forecasting times (after the model is trained) and lower overall compute requirements (including training) for a given resolution. However, this model is very data hungry: for their paper the authors used a training data set that is tens of terabytes in size. We aim to improve on such a data-driven model by including aspects of NWP into it, namely information on the physical laws which govern the weather. These laws can be integrated into the existing parametric infrastructure by introducing constraints in the loss function of the model that are grounded in domain knowledge (e.g., the conservation of mass, momentum and energy). As a proof of concept, we produce a baseline assessment of FourCastNet's performance on a small domain and examine the expected training impact and performance of our physics-informed model. We thus describe the potential of a full-scale implementation of our approach to address limitations of the purely data-driven model. With the recent advancements in the field of scientific machine learning, the timing is right for weather and climate models which can leverage both our deep understanding of physical laws and the immense amount of atmospheric data that is available.

1 Motivation

The parametric approach to weather modeling is well-motivated: massive databases containing atmospheric and weather data across the world and spanning decades are now freely, publicly available. For example, the ERA5 data set [1]. FourCastNet is a recently published, data-driven global weather prediction model with demonstrated advantages over contemporary NWP-based models [2]. However, there are limitations to such models. While the amount of available data is a boon to predictive models, training such models becomes increasingly unwieldy as downloads on the order of tens of terabytes are required. Furthermore, purely data-driven models are currently restricted to local and global weather prediction and cannot yet forecast climactic trends, such as global warming [2].

These are all problems with can be addressed using the same approach: introducing domain knowledge into the training of the model. Such methodology has found massive success in physics-informed neural networks (PINNs) [3]. Similarly, several other approaches, such as variational autoencoding neural operators [4] and DeepONets [5], have had success describing non-linear systems in terms of latent data spaces or by learning various classes of partial different equations (PDEs), which describe the physics of the weather. In fact, one part of the model used in FourCastNet, Fourier neural operators, has recently been shown to be effective at learning and solving PDEs [6]. We are thus presented with exactly the tools necessary to leverage both decades of data and hundreds of years of scientific understanding for building the next generation of weather-prediction models, the work we propose here represents the first several steps in this direction.

Our research question can be summarized as: can we improve the resolution, accuracy, and/or data requirements of a cutting-edge data-driven weather prediction model by integrating the underlying physics into its training? If the additional computation involved to achieve this occurs during training, and considering that the popular ensemble approach depends on prediction time, can this be done at little effective cost? Specifically, we aim to double the achievable resolution (4x the output grid points), increase the model accuracy by an order of magnitude, decrease the model data requirements by an order of magnitude, or some combination of these improvements. While the Fourier neural transformer used by FourCastNet is already discretization-invariant, by lowering data requirements we will also improve the generalization of the model into regions of sparse, imprecise, or non-uniform data coverage (spatially and temporally). Our augmentation of the model will not increase the prediction time of the model by more than 10%, with the increase in training time being no more than 100%. We will achieve this by integrating parts of the primitive equations into the loss function of FourCastNet.

The work we propose will answer our research question as well as serve as groundwork for future scientific machine learning-based weather and climate models. Such models will enable the synthesis of advancements in machine learning and atmospheric science, with wide-reaching benefits. Within these fields, application of novel methods can motivate future innovation and research. Beyond these fields, weather prediction finds countless applications across industries and down to the level of the individual.

2 Data and Modeling Problem

2.1 Data

ERA5 (from "The ERA5 global analysis") which consists of hourly estimates of several atmospheric variables at a latitude and longitude resolution of 0.25° from the surface of the earth to roughly 100 km altitude from 1979 to the present day and obtained through reanalysis of data at a few key locations [1]. This data allows for a plethora of models to be trained to predict the subsequent states. They also incorporate a form of data assimilation that allows for interpolated states to closely match future observations.

2.2 Modeling Problem

The modeling problem is exceedingly complex, given the many years forecasting groups have been working on a numerical solution. Though these have been improving gradually, they involve a heavy amount of computational resources to solve. Some of the factors we are considering can be found below:

Notation

- x: the x coordinate
- y: the y coordinate
- p: the pressure coordinate
- u: the wind velocity at x direction
- v: the wind velocity at y direction
- ω: the wind velocity at p direction
- $\vec{v}_H = (u, v, 0)$: the 'horizontal' component of wind velocity
- ρ: the mass density
- T: the temperature
- Φ: the geopotential
- f: the Coriolis force and is equal to $2\Omega\sin(\phi)$, where Ω is the angular rotation rate of the Earth ($\sim 7.3 \times 10^{-5}$ rad/s), and ϕ is the latitude
- c_p: the specific heat capacity (approximately constant over a certain pressure range, $\sim 29.1\ J/(\text{mol}\cdot\text{K})$
- R: the specific gas constant $\sim 8.3\ J/(\text{mol}\cdot\text{K})$
- J: the heat flow per unit time per unit mass

PDEs We consider the primitive equations in pressure coordinates. There are 5 unknowns as real-valued variables: u, v, ω, T, Φ along with the following 5 equations. However, we may not incorporate all of these equations into the loss functions due to the unavailability of certain data (e.g., the heat flow per unit time per unit mass J).

- Conservation of mass

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial \omega}{\partial p} = 0, \tag{1}$$

- Conservation of momentum

$$\begin{cases} \frac{\mathrm{D}u}{\mathrm{D}t} - fv = -\frac{\partial \Phi}{\partial x}, \\ \frac{\mathrm{D}v}{\mathrm{D}t} + fu = -\frac{\partial \Phi}{\partial y}, \end{cases} \tag{2}$$

  where

$$\frac{\mathrm{D}}{\mathrm{D}t} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + \omega\frac{\partial}{\partial p}$$

- Conservation of energy

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} - S_p\omega = \frac{J}{c_p}, \tag{3}$$

  where

$$S_p \equiv \frac{1}{c_p}\frac{RT}{p} - \frac{\partial T}{\partial p},$$

Table 1: FourCastNet Baseline Results by Varying Amount of Training Data. Note: Currently in terms of number of years, will consider flipping the axes when we get results for our experiments.

Years of Training Data	U-Direction Valid RMSE	V-Direction Valid RMSE
1	3.295	3.491
4	2.903	3.047
9	2.427	2.485

- Diagnostic equation of state for an ideal gas

$$\frac{\partial \Phi}{\partial p} = -\frac{RT}{p}, \tag{4}$$

3 Prior Work

Our parametric model will be based on the **FourCastNet** model. We aditionally considered the Pangu-Weather model at an early stage. There are additional models, but for the most part they have low resolution and these two are the main ones demonstrating to have high-resolution capabilities (with Pangu-Weather beating out the scientific model). Both of them use a transformer based architecture with different encoding schemes; FourCastNet with Fourier neural operators and Pangu-Weather with a earth-specific block layer, where the positional encoding is given by the position upon the Earth's spherical layout. We choose to evaluate on the FourCastNet model compared to the Pangu-Weather model due to its ease of accessibility (Pangu-Weather model is not completely open-sourced) and speed of training.

FourCastNet utilizes surface variables of the zonal wind velocity 10m from the surface and temperature at 2m from the surface. In addition, they incorporate temperature, zonal velocity, meridional velocity, geopotential, and relative humidity at different vertical levels determined by atmospheric pressure. Finally, they include total column water vapor.

4 Baseline

There are two options for this input format. Either we could use global inputs at a lower resolution, or more local inputs but at higher resolution. Global inputs would circumvent the need for boundary values, but local inputs would be better in terms of more specific predictive power.

For our baseline, we run the FourCastNet model on a subset of the full ERA5 dataset with a resolution of 2.5° for training and testing. The dataset is split into years 2007-2015 (with corresponding subsets for various years in training), 2016 for validation, and 2017 for out of sample testing.

Evaluation is done on the corresponding wind velocities at 10m from the surface. In FourCastNet, the authors predict both wind velocity along with precipitation, but they integrate additional considerations for precipitation that we will not explore in this work due to the purpose of this work as being a initial proposal. They also adapt a purely data-driven approach and as a result cannot enforce physical real world constraints upon the predictions of the model. Wind speed is given at various areas around the globe by means of reanalysis of trends which are measured at certain points across the globe. Our goal is multidimensional, but the primary prediction (wind speed prediction in the pure sense) has already been achieved, our main contribution would be to enforce physical constraints in order to have the model generalize better with less data. To really assess our output, a full-scale high resolution output is not necessarily required; examining our code's performance and data requirements and simply ensuring that that the output is not negatively affected is sufficient.

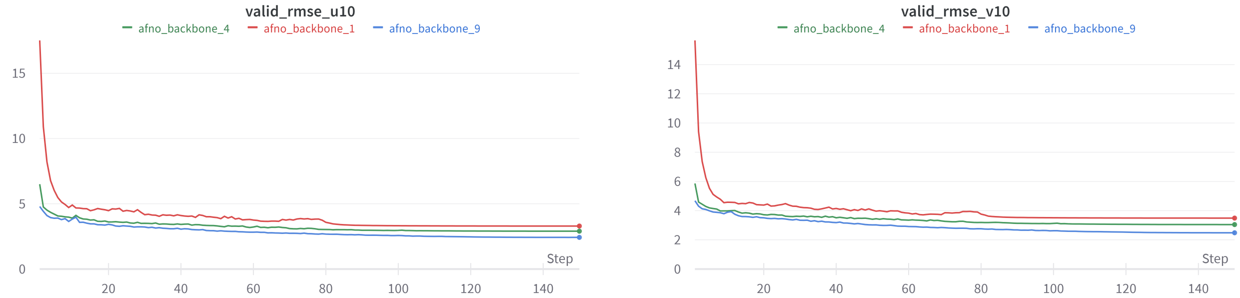
In Table 1, we show initial baseline results for the FourCastNet model and their corresponding u and v wind speed RMSE errors across differing amounts of training data. We trained each model for 150 epochs with a decaying learning rate. The full training curve is also pictured in Figure 1.

5 Expected Results

To incorporate the laws of physics into the parametric model, we consider using the physics-informed FNO [7] but replacing FNO with FourCastNet for comparison with the baseline. We consider minimizing the composite loss function

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{data}}(\theta) + \sum_{i=1}^{N_{eqns}} \lambda_i \mathcal{L}_{\text{physics},i}(\theta),$$

where $\mathcal{L}_{data}(\theta)$ is the mean square error between the ground truth and the outputs of the model, and $\mathcal{L}_{\text{physics},i}(\theta)$ is the mean square residual of equation i, which serves as the constraints of the outputs in the



(a) U10 RMSE

(b) V10 RMSE

Figure 1: RMSE Plots for Differing Amounts of Training Data

same manner as PINNs [3]. For now, we consider incorporating the equations (1) and (2). The loss weights λ_i could be fixed during training, but we can further apply more advanced algorithms to adaptively change them. Examples include using learning rate annealing algorithms [8], Neural Tangent Kernel (NTK) [9], the penalty method and augmented Lagrangian method [10].

Equation (1) produces one additional mean square error loss term:

$$\mathcal{L}_{\text{physics},1}(\theta) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{\partial \hat{u}_\theta}{\partial x}(I^{(i)}, P^{(j)}) + \frac{\partial \hat{v}_\theta}{\partial y}(I^{(i)}, P^{(j)}) + \frac{\partial \hat{\omega}_\theta}{\partial p}(I^{(i)}, P^{(j)})\right)^2,$$

where the loss is calculated for M feature inputs $\{I^{(i)}\}$ and N query points $\{P^{(i)} = (x^{(i)}, y^{(i)}, p^{(i)})\}$ on the spatial-time domain. Similarly, Equations (2) introduce two loss terms:

$$\mathcal{L}_{\text{physics},2a}(\theta) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{\mathrm{D}\hat{u}_\theta}{\mathrm{Dt}}(I^{(i)}, P^{(j)}) - f\hat{v}_\theta(I^{(i)}, P^{(j)}) + \frac{\partial \hat{\Phi}_\theta}{\partial x}(I^{(i)}, P^{(j)})\right)^2,$$

$$\mathcal{L}_{\text{physics},2b}(\theta) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{\mathrm{D}\hat{v}_\theta}{\mathrm{Dt}}(I^{(i)}, P^{(j)}) + f\hat{u}_\theta(I^{(i)}, P^{(j)}) + \frac{\partial \hat{\Phi}_\theta}{\partial y}(I^{(i)}, P^{(j)})\right)^2,$$

Thus, the final loss function is

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{data}}(\theta) + \lambda_1\mathcal{L}_{\text{physics},1}(\theta) + \lambda_{2a}\mathcal{L}_{\text{physics},2a}(\theta) + \lambda_{2b}\mathcal{L}_{\text{physics},2b}(\theta) \qquad .$$

In a full implementation the remaining equations (3) and (4) will also contribute loss terms, with the potential for an added term in the energy equation to represent long-term climactic changes such as global warming. By adding these additional, physics-based loss contributions we ensure that the model respects the laws of physics, in particular when interpolating or predicting between data points in space and/or time. As stated, we expect that this augmentation will allow for doubling of the achievable prediction resolution in each spatial dimension, and/or decrease the data demands of the model by at least an order of magnitude. This is due to the improved generalization of the model between data points (spatially and temporally), as well as in regions where data may be sparse, irregular, or imprecise. Furthermore, we also anticipate a potential increase in prediction accuracy for a given resolution and data set by an order of magnitude, due to the augmented loss disincentivizing portions of the loss parameter space which may be numerically valid (according to a ground-truth MSE loss) but physically implausible.

6 Risks and Potential Mitigation

The primary risk in the proposed work is an increase in the computation time required, primarily for training but potentially also for prediction. Adding physical constraints increases computational budget when computing the backpropagation of the neural network. However, we expect that we will use **less training data** to achieve the same level of generalization accuracy. Therefore, compared to purely data-driven models, we expect that our model will cost the same order of computational budget as FourCastNet alone. Moreover, physical constrains on the loss function will improve generalization accuracy in the limited data regime, which helps to offset the potential risks (e.g., the increased computational complexity).

There can be a risk of people trusting a weather model too much when it's good, which can lead to bad decision making for extreme weather conditions. As a result, an ensemble approach, which considers multiple simulations along with some stochasticity injected into the data, can be used to introduce some variability in the predictions as compared to a single deterministic outcome. In practice, weather conditions are usually built on such an ensemble of model predictions, providing some level of statistical uncertainty in predictions based in part on the uncertainty in the input measurements. These uncertainties can be communicated to the public (and often are, e.g., for predicted hurricane paths) to prevent undue confidence in model predictions. Since we do not expect our augmentation of FourCastNet to significantly increase the prediction time of the model, and because we expect some level of improvement to the generalization of the model (including within the space of measurement uncertainty), this ensemble approach to making statistical predictions can still be efficiently performed.

7 Code Repository

The code is available at https://github.com/ps789/FourCastNet. This is a fork of FourCastNet, with our modifications mostly residing in physics_loss.py.

References

[1] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thépaut, "Era5 hourly data on single levels from 1940 to present," *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 2023.

[2] J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, *et al.*, "Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators," *arXiv preprint arXiv:2202.11214*, 2022.

[3] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[4] J. H. Seidman, G. Kissas, G. J. Pappas, and P. Perdikaris, "Variational autoencoding neural operators," 2023.

[5] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via deeponet based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, pp. 218–229, Mar 2021.

[6] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," in *International Conference on Learning Representations*, 2021.

[7] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, "Physics-informed neural operator for learning partial differential equations," *arXiv preprint arXiv:2111.03794*, 2021.

[8] S. Wang, Y. Teng, and P. Perdikaris, "Understanding and mitigating gradient flow pathologies in physics-informed neural networks," *SIAM Journal on Scientific Computing*, vol. 43, no. 5, pp. A3055–A3081, 2021.

[9] S. Wang, X. Yu, and P. Perdikaris, "When and why pinns fail to train: A neural tangent kernel perspective," *Journal of Computational Physics*, vol. 449, p. 110768, 2022.

[10] L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, and S. G. Johnson, "Physics-informed neural networks with hard constraints for inverse design," *SIAM Journal on Scientific Computing*, vol. 43, no. 6, pp. B1105–B1132, 2021.