Towards a Dynamic Data Driven AI Regional Weather Forecast Model

Sophia Hamer, Jennifer Sleeman, Milton Halem

University of Maryland Baltimore County Baltimore MD 21250 USA
{chamer1,jsleem1,halem}@umbc.edu

Abstract. The advent of long-term reanalysis datasets such as ECMWF ERA 4/5 has enabled the development of AI-driven machine learning models for weather forecasting. The major benefit of AI as an approach is its ability to reduce computational forecast time from tens of hours to tens of seconds, thereby enabling a variety of new applications ranging from extreme regional weather event forecasting to first responder aid for wildfires, severe storms, floods, oil spills, tornadoes, and other extreme events in real time. Today, several operational weather forecast centers are evaluating these models as complements or alternatives to their existing models. However, similar efforts in applying AI/ML approaches to mesoscale weather forecasting have lagged behind due to the lack of a reanalysis of current operational regional weather forecast models. Recently, the ECMWF made publicly available Copernicus European Regional ReAnalysis (CERRA) at spatial resolutions of 11km (0.10) and 5.5km (0.050) from 1984 to the present. We present the first demonstration of a successful AI regional forecast at 5.5 km spatial resolution employing the Nvidia FourCastNet (FCN) model with its Adaptive Fourier Neural Operator (AFNO) and transformer self-attention modeling approach. We describe the training of a regional FourCastNet model in the NASA Center for Climate Studies (NCCS) Adapt cluster at the Goddard Space Flight Center using five years of CERRA reanalysis data at 3-hour intervals for five variables at four pressure levels. We show the RMSE forecast errors of a 5.5km implementation trained on five years of data improved for all variables but one over a forecast trained on three. We also devise a nesting scheme wherein our regional model is boundary forced by a global forecast. We find that our model improves on the performance of the global model over its region in all but one atmospheric variable.
Keywords: Dynamic Data Driven Learning, Regional Reanalysis, AI forecast, FourCastNet, CERRA data Sets

1 Introduction

The availability of the ECMWF ERA5 reanalysis dataset has led to a revolution in the development of AI based global models. A reanalysis data set combines long-term records of conventional ground and spaceborne observations with advanced data assimilation forecast models to form a complete global and physi-

cally consistent dataset representing the best estimate of the state of the atmosphere. Recent AI-based global weather forecast models, such as Huawei Pangu [14], Nvidia FourCastNet [10], Google GraphCast [6], and Microsoft ClimaX [9], show forecast ability comparable to operational weather forecast models, but with a significant gain in speed and a large reduction in high-performance computing requirements. For example, a traditional global weather forecast model with a spatial resolution of 1 km that takes 12 hours to simulate a 10-day forecast on today's most powerful supercomputers would take only a few minutes for an AI-based inference forecast using a trained model. This increase in speed enables the integration of regional forecast models into a digital twin for real-time weather and climate applications. For example, a wildfire digital twin can provide useful guidance to firefighters in the field or to save lives by early issuance of orders to families to evacuate their homes.

Unlike global forecasting, AI-based mesoscale weather forecasting has been less explored, as there has been a lack of reanalysis data for operational regional weather forecast models. An exception is that on 26 March 2024, Huawei announced their first 3km regional five-day forecast [8]. Unfortunately, this announcement has not been followed by public access to the publication of such models or the accuracy of their forecasts. One reason why this problem area is less explored in terms of AI research is the lack of a long-term high-resolution mesoscale reanalysis comparable to ERA5 [11]. Additionally, unlike physics-based regional models that work everywhere given the local topography, a unique reanalysis would be needed for each geographic domain of application interest. Fortunately, in 2022-2023, the ECMWF made publicly available a 40-year reanalysis for a European domain called the Copernicus European Regional ReAnalysis (CERRA) data sets at resolutions of 11km (0.10) and 5.5km (0.050) from 1984 to the present. Using an implementation of the Nvidia FourCastNet (FCN) model [10], we developed a new AI regional forecast model trained on five years of CERRA reanalysis data at 3-hour intervals for five variables in four pressure levels.

Due to the relatively unexplored nature of regional AI weather forecasting models, we must determine what adaptations are needed to achieve performance comparable to global models over each region, as well as explore different methods of allowing the global model to influence the regional model in order to allow global weather effects to influence the region as part of a nesting scheme.

2 Relation to Dynamic Data Driven Applications Systems

The Dynamic Data Driven Applications Systems paradigm describes the integration of instrumentation data and application models together into a feedback control loop, which can then be used to control the instrumentation or improve the accuracy of the model of the physical system [2]. DDDAS has found uses particularly in the modeling and structure of digital twins [7] [12]. Traditional mesoscale forecasts rely on slow and computationally intense simulations, mak-

ing it difficult for them to be guided by dynamic data input. Our work's major contribution is the large reduction in computational forecast time and resources for a regional weather model, which allows integration into a dynamic control system within a DDDAS paradigm. Faster forecasts also enable new applications within this paradigm, such as live monitoring and simulation of extreme weather events such as wildfires. In turn, this can be used to steer instrumentation to collect and measure data relevant to generating a more accurate model of the event in order to (for instance) better inform firefighters and emergency services. Our model fits within the data assimilation loop of the DDDAS method, as its quick simulation time allows for faster updating of sensor instrumentation and integration of sensor data. Figure 1 illustrates a basic version of this DDDAS system.

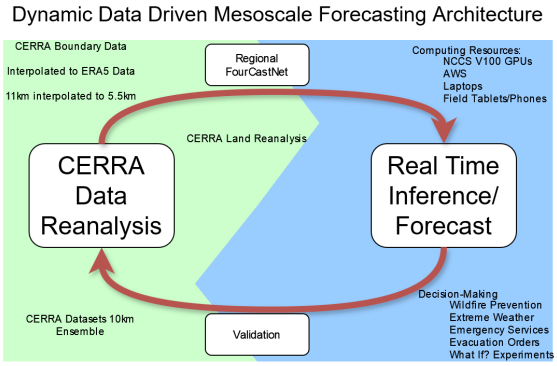

Fig. 1: Illustration of a simple DDDAS system integrating our regional FourCastNet mesoscale model into CERRA reanalysis. This can itself be used in a larger DDDAS system to drive instrumentation based off this reanalysis.

3 Background

There are two general types of weather forecasting models: global models at 30km spatial resolution and mesoscale models at 3km resolution (both are used operationally for different applications). The use of global weather models for regional forecasting at 3km scale is a 10x factor in spatial resolution and a 100x temporal computing constraint even with today's fastest supercomputers, which limits such use to non-real time exploratory research. Mesoscale weather models can perform forecasts at these resolutions and higher by defining geographic domains at 1/8 or less the size of the global domain and embedding nested domains of higher resolution covering even smaller geographic areas. For instance, we can embed a higher resolution model of the continental United States within a lower resolution global model. Nested geographic areas add the complexity of having to specify interactive one-way or two-way boundary conditions.

Mesoscale Weather Forecast models are governed by the same dynamical equations that govern the atmospheric motions of Global Weather Forecast models [13, 1]. Although regional models employ many of the same physical parameterizations as global models such as radiation, convection, and microphysics, they also include processes such as vegetation surface scale modules, soil type and depth, local topography for wildfires, surface runoff, and even local chemical effects[4]. As a result of trade-offs in geographic coverage for increased spatial resolution, regional models can numerically simulate atmospheric motions on domains that are an order of magnitude smaller than global models. This increased spatial resolution allows regional models to address local processes taking place on shorter time scales such as boundary layer processes, wildfire models, air quality, snow cover and melt, and mountain effects, etc. However, higher spatial and temporal resolutions come at the cost of increased compute time and resources comparable to those of lower resolution global forecast models. Thus, while regional models with very high resolution can realistically simulate more local phenomenon such as wildfire spread and the dispersion of smoke from plumes, they cannot provide real-time decision-making firefighting guidance, a desired property of wildfire digital twin (for example).

4 A CERRA Data Driven FourCastNet for Regional Predictions

We chose to employ the global AI model FourCastNet (FCN) [10] developed by Nvidia, but limited to the domain of the CERRA data described above for continental Europe. FCN is a model based on the use of Adaptive Fourier Neural Operators (AFNO) by J. Gubias [3]. Each spatial variable at a given pressure level in the training data set is stored as an image of the spatial domain of the dataset for each timestep. First, FCN partitions the image into patches of dimension (h,w). Each patch is then embedded as a large vector along with the other variables and sent to an AFNO. The AFNO then transforms this input with a Fast Fourier Transform (FFT) followed by a linear/softmax layer. In essence, the model learns the change in the Fourier coefficients from this timestep to the next. The coefficients are then sent through an inverse FFT (IFFT) and then mixed back with the original input. This process is repeated some number of times (eight by default) and then sent to a final linear layer to decode this back into an image. See Figure 2 for an illustration of this process.

A model trained in this way would only be able to take in input over the CERRA domain and would not know anything about the state of the atmosphere outside it. Numerical models tackle this issue by employing nesting: a low-resolution global forecast is run first, then the results of that forecast are used to generate boundary conditions for a higher-resolution regional forecast. This allows the global atmospheric state to drive the regional predictions.

In the case of our CERRA model, we used the results of a run of Nvidia's original global model trained with five years of ERA5 data as boundary conditions for our regional model. We selected the same variables and years as our CERRA

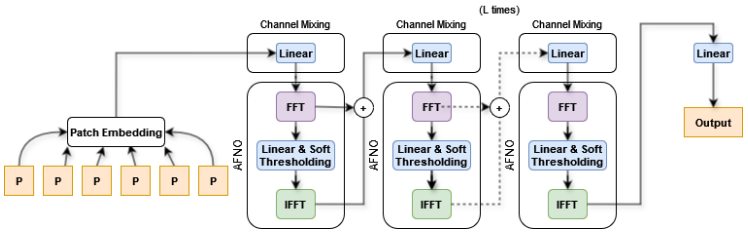

Fig. 2: An illustration of the FourCastNet model. The input images are broken down into patches which are then embedded and sent to a series of AFNOs, then decoded back into images at the end.

model for the training data for this model. To implement nesting, we interpolate from the low-resolution 30km ERA5 grid to the higher-resolution 5.5km CERRA grid. We then run inference using our regional CERRA-trained model, but after each time step completes and errors and predictions are calculated, we integrate all pixels that are within N pixels of the edge (either height- or width-wise) on each variable's image representation with the interpolated results from the ERA5-based model's inference run, then resume inference as normal. We utilize a relatively basic data assimilation technique to drive the regional variables: for each variable, we simply replace all the relevant pixels with the corresponding pixel on the interpolated global forecast.

5 Training Regional AI FourCastNet on CERRA Reanalysis

We first modified FCN to support a regional domain. Although the version of FCN we used lacks periodicity, it still required modification to take in data of different resolutions and shapes to the original ERA5 dataset that it was initially trained with. Since FCN makes no assumptions about the nature of each channel/variable, we were free to choose any variables from the CERRA reanalysis data. We selected temperature, geopotential, wind speed and relative humidity to match the variables in the original ERA5 dataset, and selected 50, 500, 850, and 1000 hPa as our pressure levels for the same reason.

In particular, the specific CERRA dataset we used did not have surface-level (2m) data for these variables, though future experiments will incorporate other datasets that do. We also did not omit certain variables from certain pressure levels, as Nvidia's ERA5 dataset did - every pressure level in our model contains every atmospheric variable. Finally, our CERRA dataset did not provide total column water vapor, mean sea level pressure, and surface pressure, as does the ERA5 dataset. This is also something that we will incorporate into future experiments.

The CERRA data we used had a resolution of (1069,1069), however, for ease of patching, we used multiples of eight and achieved this by truncating

the bottom of rightmost pixels of the image to achieve a size of (1024,1024) - a spatial resolution of 5.5km (0.050). With 20 channels, this model has a tensor input of (20, 1024, 1024). The CERRA dataset also has a temporal resolution of three hours per time step rather than ERA5's six hours.

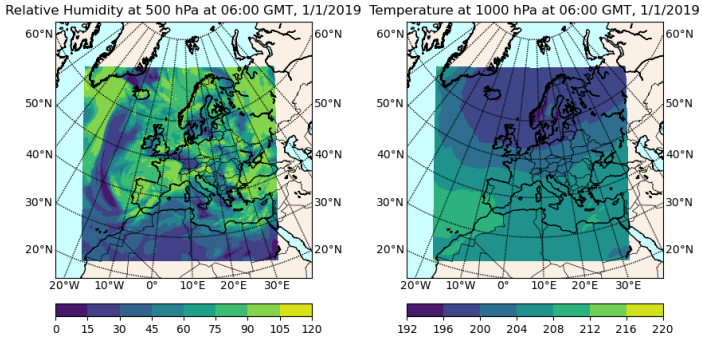

Fig. 3: Plots of forecasted temperature (in Kelvin) at 1000 hPa and relative humidity (in kg/kg) at 500 hPa at 0600 GMT on January 1, 2018 (two timesteps into the forecast) with five years of training data.

Although CERRA reanalysis data sets are now available for 40 years, the computational resources needed to store and train 150 epochs of a 40-year 5.5 km data set are still quite high. However, in an earlier study presented at the 2023 American Geophysical Union Annual Meeting[5], we showed that two years of training with only 50 epochs in the ERA5 dataset approximates Nvidia's complete FCN model forecast for 2018 (which used 40 years of data at 150 epochs) reasonably well with less than a 20% RMSE error for one to two days. This suggested that fewer training years for our regional model as well as fewer epochs could be used to overcome hardware challenges while still maintaining reasonable skill.

We conducted three relevant experiments with one, three, and five years of available CERRA training data, respectively, starting in 2013 and ending in 2017. The one-year experiment was trained on just 2013, three years was trained on 2013-2015, and five years was trained on 2013-2017. 2018 was used as the testing year, just as in Nvidia's initial FCN training on the ERA5 dataset. The one-year training data experiment was used to test the execution of the regional FCN model and was trained for the full 150 epochs.

Our two primary experiments were the three-year and five-year training data runs, where we collected both images of completed forecasts and RMSE plots for each variable for those forecasts. Both experiments were left to train for a reduced number of epochs: 30 for three years and 60 for five years. Using trained models, we produced five-day forecasts starting at 00:00 GMT on January 1, 2018. Examples of forecast images for the five-year run can be found in Figure 3, and examples of RMSE plots are found in Figure 4.

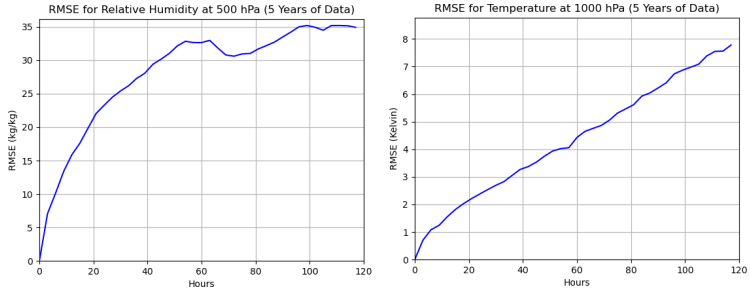

Fig. 4: Plots of the RMSE of temperature (in Kelvin) at 1000 hPa and relative humidity (in kg/kg) at 500 hPa at 0600 GMT on January 1, 2018 (two timesteps into the forecast) given five years of training data

We used the computing resources made available on the NASA NCCS system at Goddard Space Flight Center. We were limited to four Nvidia V100s each with approximately 32GB of VRAM, which took approximately four days to fully train each experiment. For comparison, NVidia's initial training of FCN (40 years of data, 150 epochs) on the ERA5 dataset was done across 64 A100s, each with 80GB of VRAM, taking approximately 16 hours.

We now describe the nesting experiments conducted. We varied the boundary size value N from 0 (meaning a completely unnested forecast) to 64. A special 512 boundary is included wherein the entire image is replaced with the ground truth from the global dataset interpolated to the CERRA domain. We also ran a set of experiments in which only geopotential was replaced.

6 Results of CERRA Regional AI Forecast

Running inference for a five-day forecast for both the three-year and five-year training data models took less than 10 seconds on four Nvidia V100s and less than a minute on a GeForce RTX 4080m laptop GPU. The RMSE inference errors were calculated by differencing with the reanalysis data. Other validation metrics (such as the anomaly correlation coefficient (ACC)) await the calculation of a CERRA climatology and will be included in future work. In Figures 5 - 9 we show a comparison of RMSE over time for the three-year and five-year training data experiments for all variables and pressure levels. For almost all 20 combinations of variable/pressure levels, the five-year training data experiment performs noticeably better than the three-year. These results show forecast skill for a full five-day forecast for three and five years of training data. We expect a longer number of years of training to extend the range of forecast skills.

The results of the initial nesting experiment are shown in Figures 10-14. Here, we found that our nesting scheme occasionally produced a small improvement in our forecast performance for certain variables (relative humidity at 500 hPa, wind velocity U at 850 hPa, and others) but was just as likely to be a detriment (wind velocity u at 500 hPa, wind velocity V at 500 hPa, and others). One variable

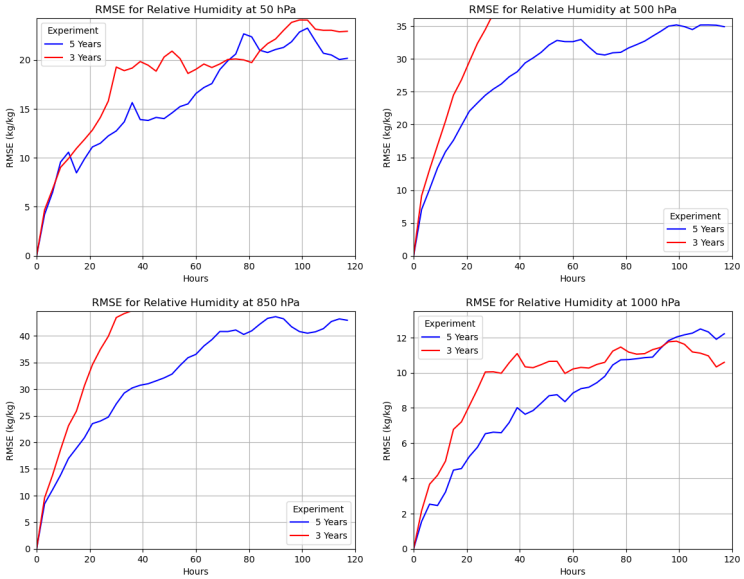

Fig. 5: Comparison of three-year and five-year experiment performance of Relative Humidity at each pressure level.

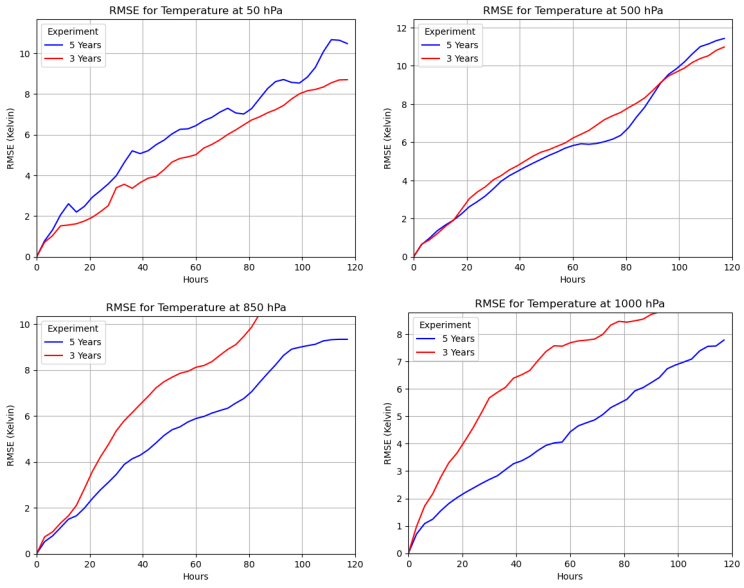

Fig. 6: Comparison of three-year and five-year experiment performance of Temperature at each pressure level.

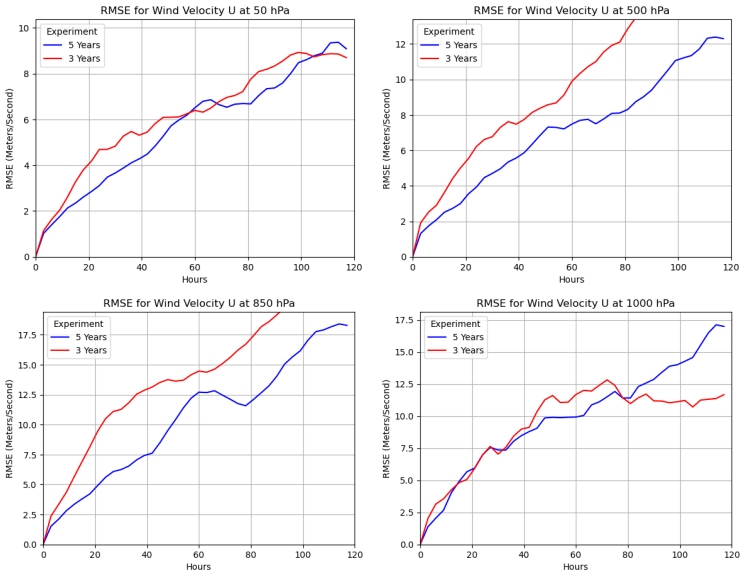

Fig. 7: Comparison of three-year and five-year experiment performance of Wind Velocity U at each pressure level.

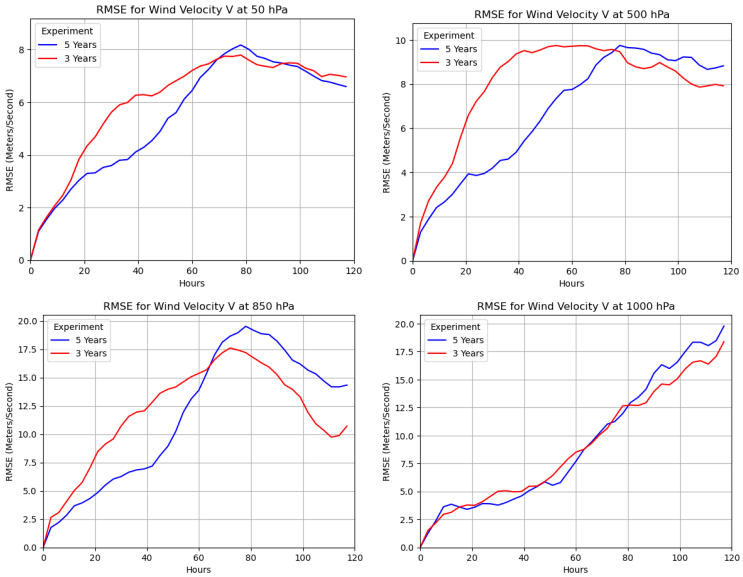

Fig. 8: Comparison of three-year and five-year experiment performance of Wind Velocity V at each pressure level.

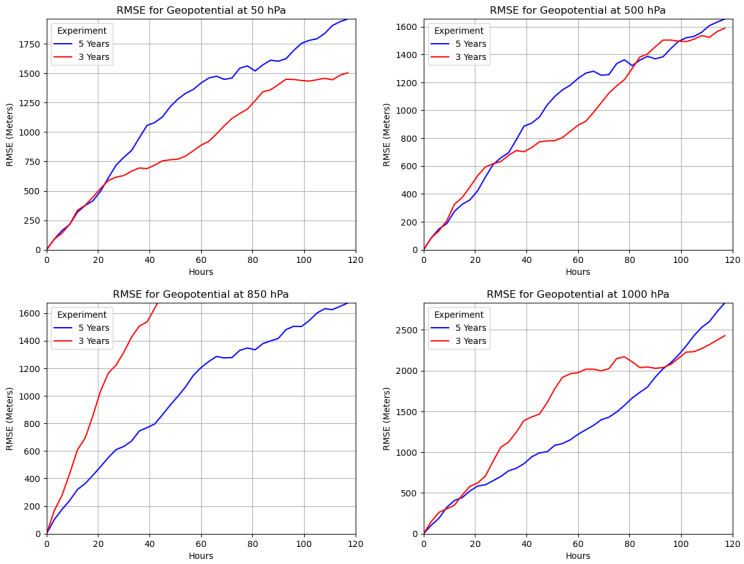

Fig. 9: Comparison of three-year and five-year experiment performance of Geopotential at each pressure level.

that stands out is geopotential: the five-year Nvidia model already predicted geopotential consistently better than our regional model, and in those cases, using nesting actually significantly improved our model's performance (though not to the point where it was performing as well as the five-year Nvidia model).

These experimental results led to an additional experiment in which only geopotential variables were replaced. The results of this experiment are shown in Figures 15-19. Note that for our only-geopotential replacement experiment, we performed better at predicting geopotential for all pressure levels than we did when we replaced all variables. This suggests that our model shows a genuine skill improvement over the global five-year model and is able to actively benefit from the boundary conditions of a global model. Importantly, we did not see a drop in performance for the other (non-geopotential) variables, and in some cases (such as wind velocity V at 1000 hPa) we saw a large drop in RMSE for many boundary sizes. In several variables, we also found that a total replacement of the geopotential (the 512 line) actually produced a large improvement in the performance of almost every other variable. This opens an avenue of future investigation, wherein we can explore leaving some atmospheric variables to be predicted exclusively by global models and simply interpolated to a regional domain for use in regional models.

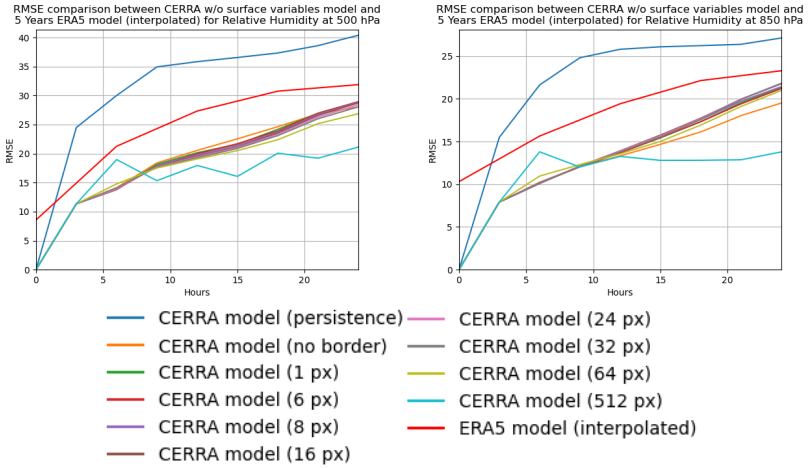

Fig. 10: Relative Humidity Comparison of the CERRA regional model and the five years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Forecast begins at 00:00 GMT January 1, 2018.

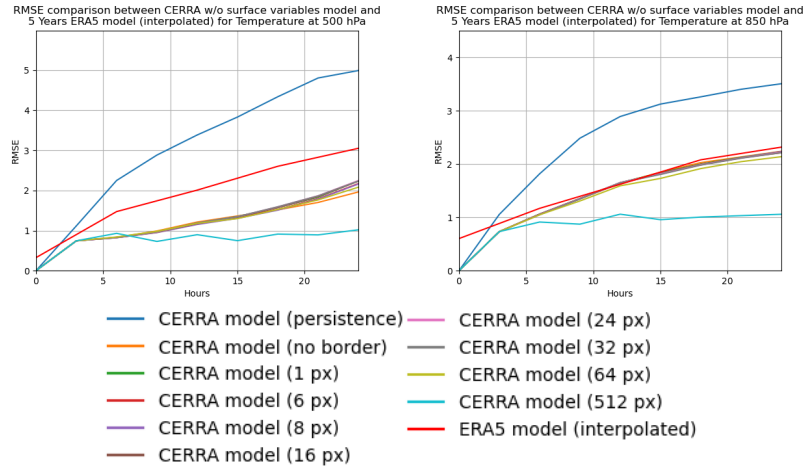

Fig. 11: Temperature Comparison of the CERRA regional model and the five years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Forecast begins at 00:00 GMT January 1, 2018.

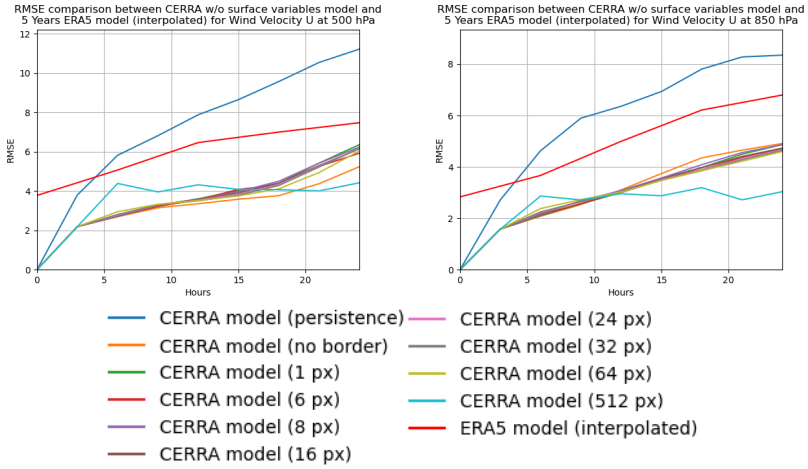

Fig. 12: Wind Velocity U Comparison of the CERRA regional model and the five years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Forecast begins at 00:00 GMT January 1, 2018.

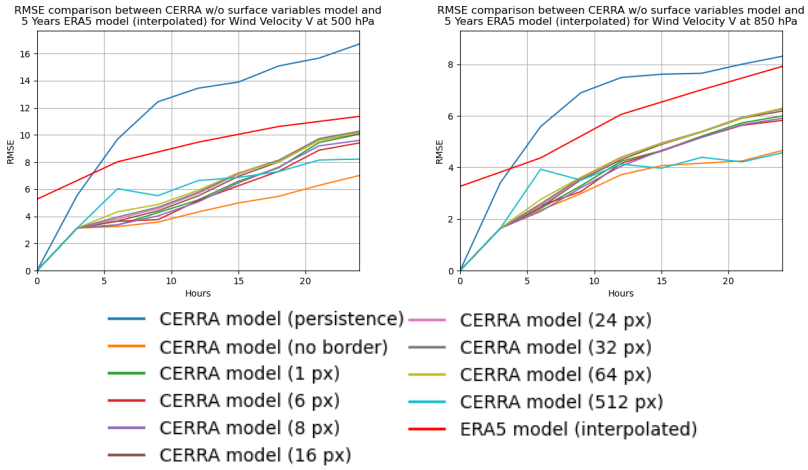

Fig. 13: Wind Velocity V Comparison of the CERRA regional model and the five years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Forecast begins at 00:00 GMT January 1, 2018.

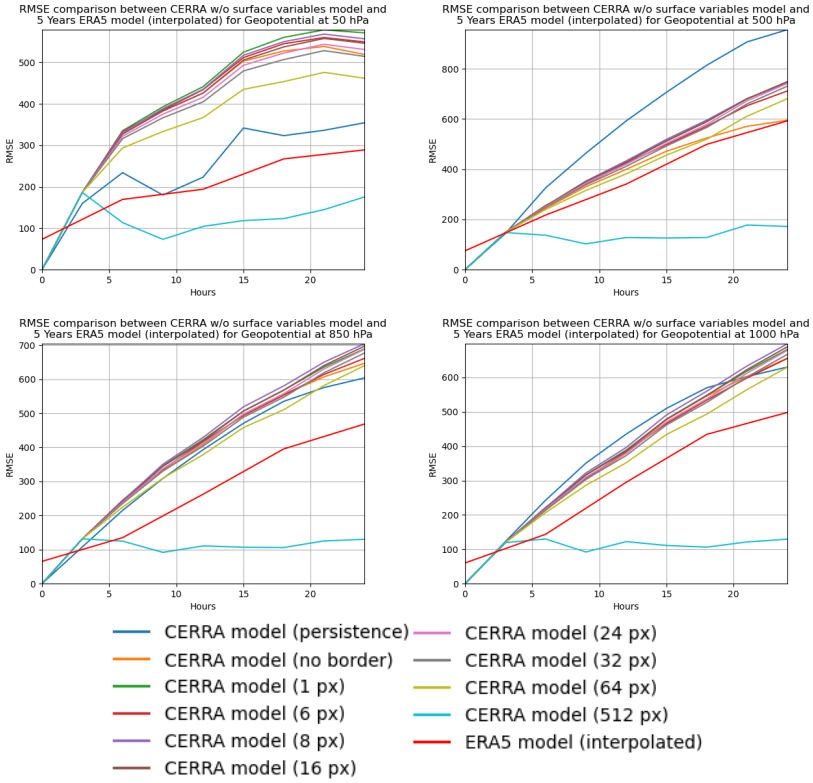

Fig. 14: Geopotential Comparison of the CERRA regional model and the five years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Forecast begins at 00:00 GMT January 1, 2018.

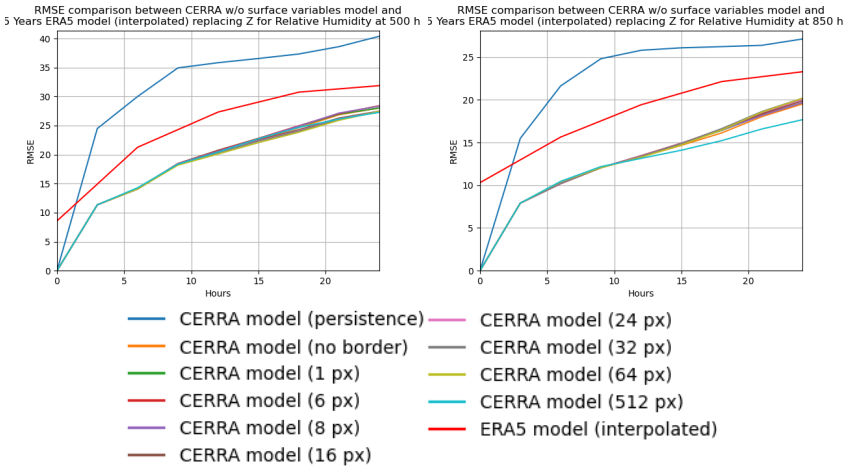

Fig. 15: Relative Humidity comparison of the above CERRA regional model and the 5 years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Here, only geopotential variables are replaced. Forecast begins at 00:00 GMT January 1, 2018.

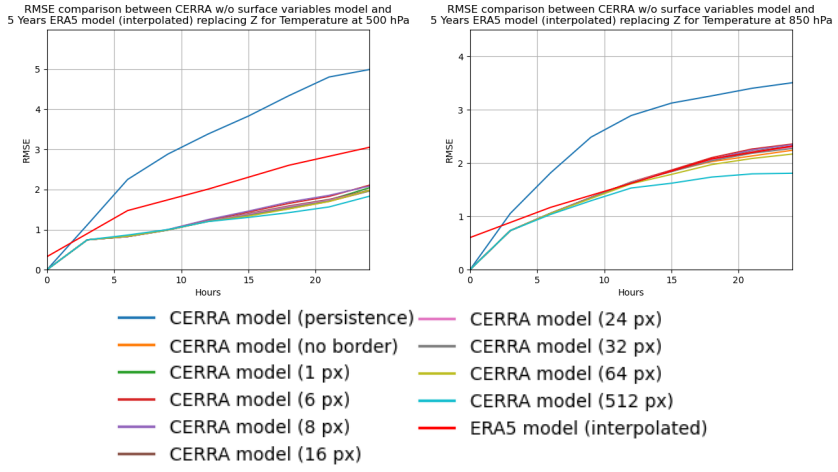

Fig. 16: Temperature comparison of the above CERRA regional model and the 5 years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Here, only geopotential variables are replaced. Forecast begins at 00:00 GMT January 1, 2018.

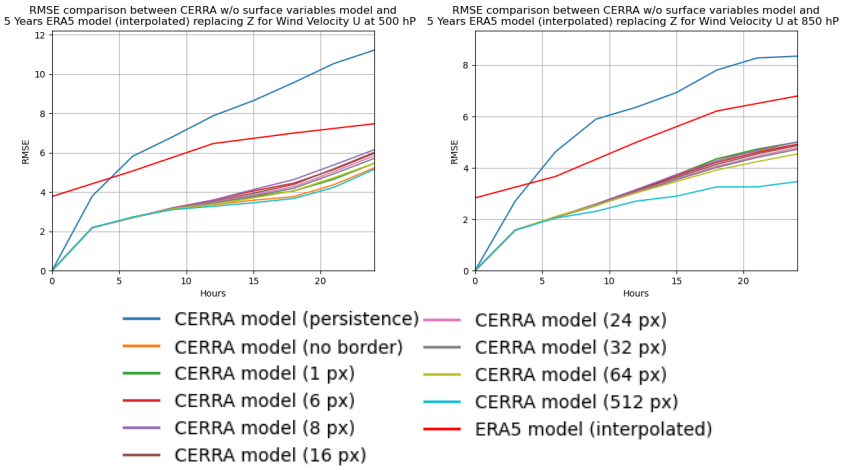

Fig. 17: Wind Velocity U comparison of the above CERRA regional model and the 5 years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Here, only geopotential variables are replaced. Forecast begins at 00:00 GMT January 1, 2018.

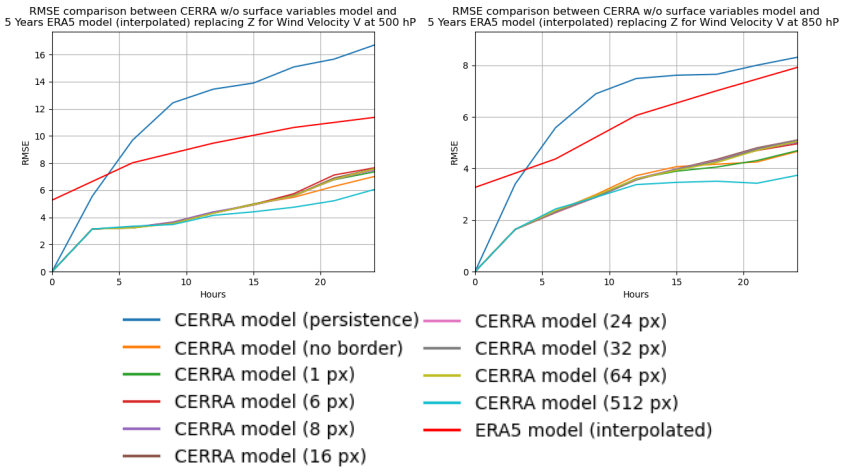

Fig. 18: Wind Velocity V comparison of the above CERRA regional model and the 5 years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Here, only geopotential variables are replaced. Forecast begins at 00:00 GMT January 1, 2018.

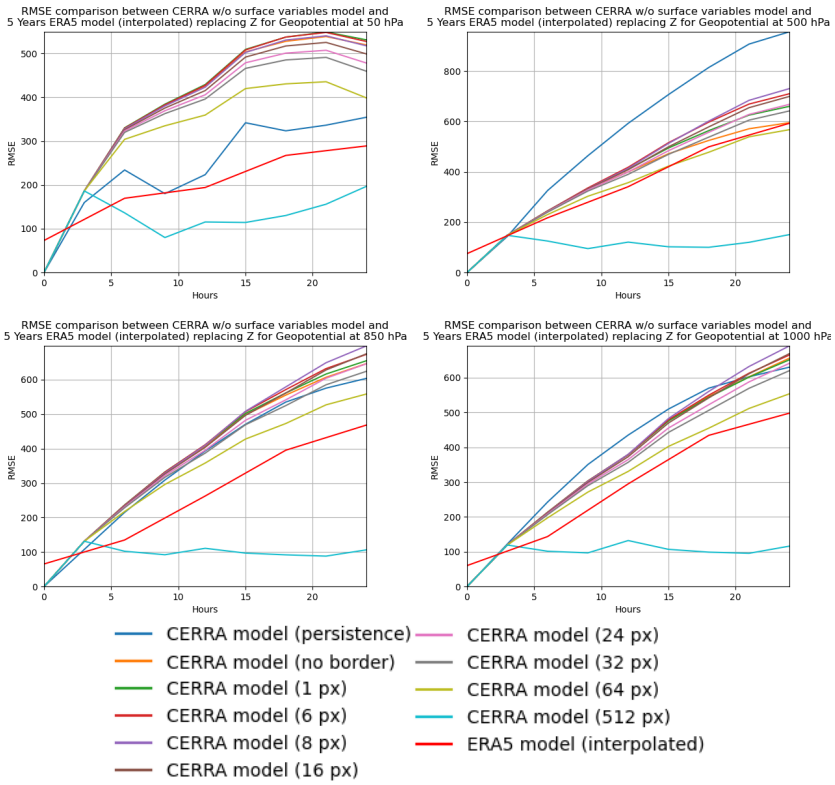

Fig. 19: Geopotential comparison of the above CERRA regional model and the 5 years of data ERA5 global model (interpolated to the CERRA domain) for different boundary sizes. For each size n, the n pixels closest to the border of an image are replaced with predictions from the global forecast before each time step. Here, only geopotential variables are replaced. Forecast begins at 00:00 GMT January 1, 2018.

7 Conclusions and Next Steps

We have taken the first step towards demonstrating the first mesoscale AI weather research forecast employing the FourCastNet model with one to five years of training data in the 40-year CERRA reanalysis data set at 5.5km spatial resolution. Even with limited training data and time, we still produce realistic 24-hour forecasts at 5.5 km. Our nesting experiments showed that our regional model can indeed benefit from having a global model modify its boundaries at each timestep under certain conditions. Although our model performs better for most variables, a global model with the same amount of training data performs better at predicting geopotential. Our experiments with switching out which

variables we actually replaced suggest that our model benefits primarily from the geopotential correction provided by the global model, and having this influence enables it to improve its performance at predicting other variables as well. Our next steps are to experiment with other assimilation schemes for boundary forcing and train the model with additional training data. We would also like to extend our analysis beyond January 2018 to better assess our model's performance, as well as test our regional nesting scheme on other regions, such as the continental US and Australia.

8 Acknowledgements

We recognize the NASA/ESTO FireSense Program Manager, Haris Riris, and his staff for their support on our grant number 80NSSC22K1405. We also acknowledge the support from the NASA CISTO office in making GSFC/NCCS computing resources available to this grant, without which these breakthrough findings would not have been possible.

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bao, J.W., Michelson, S., Grell, E.: Microphysical process comparison of three microphysics parameterization schemes in the wrf model for an idealized squall-line case study. Monthly Weather Review **147**(9), 3093–3120 (2019)
2. Darema, F., Blasch, E.P., Ravela, S., Aved, A.J.: The dynamic data driven applications systems (dddas) paradigm and emerging directions. Handbook of Dynamic Data Driven Applications Systems: Volume 2 pp. 1–51 (2023)
3. Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., Catanzaro, B.: Adaptive fourier neural operators: Efficient token mixers for transformers. arXiv preprint arXiv:2111.13587 (2021)
4. Ha, S., Kumar, R., Pfister, G., Lee, Y., Lee, D., Kim, H.M., Ryu, Y.H.: Chemical data assimilation with aqueous chemistry in wrf-chem coupled with wrfda (v4. 4.1). Journal of Advances in Modeling Earth Systems **16**(2), e2023MS003928 (2024)
5. Hamer, S., Halem, M., Sleeman, J.: Towards a regional ai-driven digital twin forecast model. AGU23 (2023)
6. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting. Science **382**(6677), 1416–1421 (2023)
7. Malik, S., Rouf, R., Mazur, K., Kontsos, A.: A dynamic data driven applications systems (dddas)-based digital twin iot framework. In: Dynamic Data Driven Applications Systems: Third International Conference, DDDAS 2020, Boston, MA, USA, October 2-4, 2020, Proceedings 3. pp. 29–36. Springer (2020)
8. Matsui, E.: Huawei Cloud introduces regional AI weather forecast Pangu model in China. https://www.huaweicentral.com/huawei-cloud-introduces-regional-ai-weather-forecast-pangu-model-in-china/ (2024), [Online; accessed 27-March-2024]
9. Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J.K., Grover, A.: Climax: A foundation model for weather and climate. arXiv preprint arXiv:2301.10343 (2023)
10. Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., et al.: Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv preprint arXiv:2202.11214 (2022)
11. Ridal, M., Bazile, E., Le Moigne, P., Randriamampianina, R., Schimanke, S., Andrae, U., Berggren, L., Brousseau, P., Dahlgren, P., Edvinsson, L., et al.: Cerra, the copernicus european regional reanalysis system. Quarterly Journal of the Royal Meteorological Society (2024)
12. Rokka Chhetri, S., Al Faruque, M.A., Rokka Chhetri, S., Al Faruque, M.A.: Dynamic data-driven digital twin modeling. Data-Driven Modeling of Cyber-Physical Systems using Side-Channel Analysis pp. 129–153 (2020)
13. Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Liu, Z., Berner, J., Wang, W., Powers, J., Duda, M., Barker, D., et al.: A description of the advanced research wrf model version 4.3; no. NCAR/TN556+ STR (2021)
14. Zeng, W., Ren, X., Su, T., Wang, H., Liao, Y., Wang, Z., Jiang, X., Yang, Z., Wang, K., Zhang, X., et al.: Pangu: Large-scale autoregressive pretrained chinese language models with auto-parallel computation. arXiv preprint arXiv:2104.12369 (2021)