Continual Learning and Adaptation In Resource-Constrained Environments (CLAIRE)

Megan Baker[1], Jennifer Sleeman[1], Christopher Ribaudo[1], Ivanka Stajner[2], Kai Wang[2], Jianping Huang[2], Ho-Chun Huang[2], Raffaele Montuoro[2], Haixia Liu[2]

[1]The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road Laurel, MD 20723
[2]National Oceanic and Atmospheric Administration, 5830 University Research Ct College Park, MD 20740
{megan.baker,jennifer.sleeman,chris.ribaudo}@jhuapl.edu

Abstract

As climate-related extreme events continue to increase and impact the world, of particular importance are the threats posed to food, agriculture, and water (FAW) systems. Deep learning could benefit FAW systems for classification of threats and for forecasting potential future events given historical patterns. However, many FAW systems are faced with operational environments that are resource-constrained, which could present challenges in deploying deep learning models. Continual learning offers a way to overcome certain deployment challenges by enabling deep learning models that are more robust to data distribution changes, without the need for GPUs or off-line training. We describe a continual learning approach to forecasting extreme air quality events developed for the National Oceanic and Atmospheric Administration to provide operational air quality guidance to the Continental United States. We describe how this deep learning model is resilient to future data distribution changes by performing curriculum learning, and how it can be deployed as a continual learner, offering better predictive performance for resource-constrained environments.

Introduction

As climate-related changes continue to impact the global society, there are increased threats to food, agriculture, and water (FAW) systems (Wiebe, Robinson, and Cattaneo 2019; Misra 2014). With sea level rise, increased flooding and drought events, and other extreme weather conditions, these systems could benefit from the use of Artificial Intelligence (AI) for forecasting potential future threats (Bahari et al. 2023; Leal Filho et al. 2022; Ben Ayed and Hanana 2021). However, these systems are often faced with operational constraints on hardware and memory that impact the type of technological advancements that can be deployed. For this reason, deep learning (DL) models that require significant resources to train or that require frequent retraining and deployment may not be an option. Of particular importance is the fact that the biggest threats are often to countries with the least amount of resources (Morton 2007). However, these resource-constrained environments should still benefit from the predictive power of DL. We propose the use of two continual learning (Wang et al. 2024) approaches that could be used as low-cost DL techniques for operationally resource-constrained environments. These approaches have the potential added benefit of being resilient to changes in input data distributions, since continual learning works to enable models that continually learn and improve. These methods could be used in operational environments without GPU resources. We describe our methods and their application for a use case with a resource-constrained operational environment, namely for National Oceanic and Atmospheric Administration (NOAA) operational air quality forecasting guidance. NOAA's operational environment is such that GPUs are not accessible, DL model deployment is limited to yearly deployment, and air quality extreme events result in distribution shifts that can result in less effective forecasting using a once-trained DL model.

Background

In this section we provide a background of the challenges of air quality forecasting and how continual learning is used for forecasting.

Air Quality Forecasting

NOAA's operational forecast guidance includes ozone and fine particulate matter (PM2.5). PM2.5 results from wildfires and is a growing concern due to the health-related effects and increased frequency of wildfires globally, in part due to climate-related changes (Jaffe et al. 2020). NOAA chose to explore DL emulation of their forecasting models to enable 3km resolution 72-hour forecasts, which is currently intractable in their environment. This requirement is due to a new adoption of the next-generation Air Quality (AQ) forecasting system coupling the Rapid Refresh Forecast System (RRFS) with the Community Multiscale Air Quality (CMAQ) Modeling System (Stajner 2022). The forecast includes chemical reactions and transport of over 183 chemical tracers. Though DL is suitable for this type of problem, an additional challenge is the deployment of such a model which must maintain skill in a resource-constrained, operational environment.

Extreme Air Quality Events and Distributional Shift While distributional shift is a commonly encountered challenge for deployed systems across many domains, extreme air quality events can result in significant anomalous behavior. As climate change continues to influence weather patterns, these extreme air quality conditions can present a considerable difficulty for forecasting. For example, the wildfire events that occurred on the West Coast in September 2020 caused a number of anomalous air quality events across many days throughout the Continental United States (CONUS). Given that air quality behavior changes both within and across years - sometimes very unpredictably - there may be challenge in maintaining system performance without requiring substantial, periodic updates to the model.

Continual Learning for Forecasting

Although significant progress has been made across many areas of machine learning, even state-of-the-art ML approaches can lack robustness in "real world," operational environments (Hadsell et al. 2020; Csurka 2017), where incoming data may differ from the original training data. Ideally, deployed systems should adapt to these differences in data distributions and maintain performance in varied contexts, that is, they should be capable of *Lifelong* or *Continual* Learning (CL) (Parisi et al. 2019; Chen and Liu 2022). Continual learning approaches are applied to many kinds of problems (e.g. classification, reinforcement learning) (Abel et al. 2024; Baker et al. 2023; Ao and Fayek 2023), but the application of these techniques to forecasting problems has been primarily limited to the energy, financial, and traffic analysis domains (Besnard and Ragot 2024). Applying CL methods to the climate domain has only recently been explored (Schillaci, Schmidt, and Miranda 2021) and may face the additional challenge of resource constraints in deployment that mandate the avoidance of overly complicated techniques or bulky models which are often employed to address issues of distribution shift.

Related Work

Previous work has explored a variety of techniques to address the challenges associated with resource constraints in model deployment. These have become especially critical as state-of-the-art models continue to increase in size (Stojnic et al. 2022). Approaches to perform DL as efficiently as possible while maintaining model performance can be grouped into categories (Menghani 2023; Ahmad et al. 2023) which include: compression techniques, learning techniques, automation, efficient architectures, and infrastructure. In this work, we focus on learning techniques, specifically CL techniques, to operate efficiently in a resource-constrained environment while attempting to be robust to potential future distribution shifts.

In particular, two methods often used to prepare for CL scenarios can be leveraged for potential impact in lower resource domains: 1) *Curriculum Learning* (Bengio et al. 2009; Faber et al. 2024), where inputs presented during model training are intentionally selected and ordered to maximize performance rather than all data being randomly shuffled before presentation, and 2) *Online Learning* (Hoi et al. 2021; Pham et al. 2022), where smaller batches of data are used to update a deployed model as they become available. These methods are of particular interest to resource-constrained environments because they do not require any model architecture modifications and could potentially be a broadly applicable technique to develop more robust and resilient FAW forecasting systems.

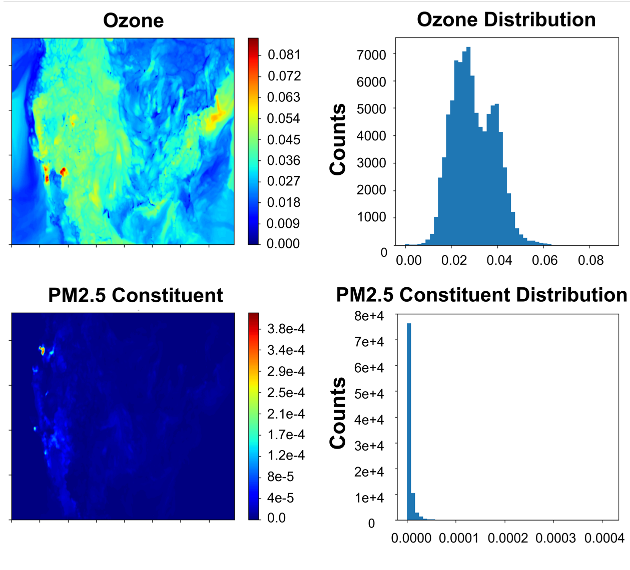


Figure 1: An example of ozone (O3) data (top left) and its distribution (top right) and of one of the PM2.5 constituents, ASVOO3J data (an organic species contributing to PM2.5) (bottom left) and its distribution (bottom right), illustrating the challenges of working with air quality data sets.

Data and Methods

This work explores initial steps toward incorporating two of the more lightweight CL techniques for a forecasting effort with particular consideration for systems deployed in a resource-constrained environment. We demonstrate that these techniques show promising results for inclusion in future efforts with similar considerations.

Air Quality Forecasting Data

To assemble our air quality forecasting dataset, we used model output at a 15km resolution from the Community Multiscale Air Quality (CMAQ) Modeling System, choosing dates starting in September 2020 in order to capture the extreme air quality events present during that time due to wildfires. The model generated outputs every 30 minutes for 30 hours, resulting in a total of 59 outputs that were generated every five days over the course of a month. These outputs consisted of 183 chemical species concentrations and meteorological variables throughout CONUS, represented by a 232x396 grid with 64 associated vertical levels. This grid was resized to a size of 128x128 to accommodate the model architecture, and only the lowest vertical level (i.e. ground level) was included. Of the 183 species, only the constituents which contribute to NOAA's air quality guidance, namely for ozone and PM2.5 were included, giving a total of (7 days) x (59 timesteps) x (94 chemical species concentrations and meteorological variables) x (128 longitude) x (128 latitude) data points. To use this data for air quality forecasting, we break up all species data from each day into sequences of 16 datapoints. This gives four hours of historical data to input to the model, and the outputs to be predicted are the computed differences from each of the four hours into the future. Only sequences of 16 data points that can be allocated without hitting a "day" data boundary are used.

Model and Training

We perform forecasting with a 3D U-Net deep learning model with six downsampling and upsampling blocks (Ronneberger, Fischer, and Brox 2015), and adapt the U-Net difference learning approach from (Chen et al. 2024) with a few key changes described below to apply to the species-agnostic air quality forecasting problem. Air quality forecasting is challenging in part because distributions of concentration levels are typically non-normal, often left or right skewed, thus presenting challenges for deep learning models even after normalization is applied. An example of ozone and a PM2.5 constituent (used to calculate PM2.5) are shown in Figure 1, along with their associated distributions. The difference method applied to this data can result in more normal distributions across constituent species.

We made some modifications to the difference learning approach described by (Chen et al. 2024). These changes include first removing the cube root transformation and instead apply a linear scaling, log transform, and min-max scale using the computed statistics of an individual day. Since this forecasting approach does not distinguish between the constituent species, the min-max scaling allows a more robust representation of a majority of the data, with only simple operations required to transform the data back to the original units. Each model had 16.7M parameters and was trained using a single NVIDIA H100 with an 80/20 train/validation split. We also use mean absolute difference to compute the element-wise loss. Training times using this configuration are described in Table 1. Evaluation was performed using data from days unseen during training and we use these results to compare performances for our experiments below.

Experimentation and Results

Curriculum Learning Experiment

Our first experiment compared a traditional machine learning approach of using all available training data to a CL approach of curriculum learning, where data is presented in a predefined way in order to maximize performance while minimizing time and resource cost. This was initially explored for forecasting by (Koenecke and Gajewar 2020), and could be especially useful since the most recent data is likely to be very relevant for future prediction; however, recent data must not be overemphasized during training in order to avoid overfitting and poor generalization.

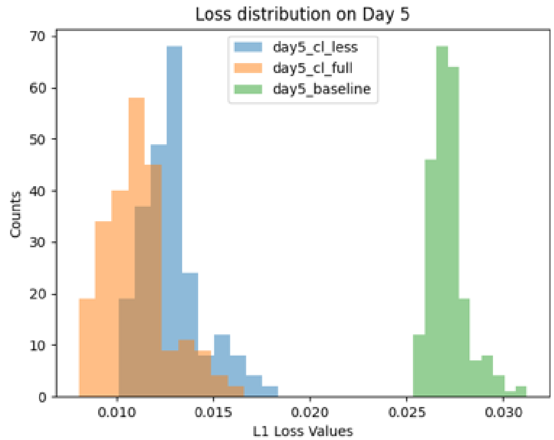
In particular, the curriculum that was chosen for this problem was to train a model using the first four days (days 0 through 3) of data for 25 epochs (*CL_full* model), and then present exclusively the fifth day (day 4) data for five epochs before testing on the sixth and seventh days (days 5 and 6). To compare the effect that the curricula had on performance, the performance of the *CL_full* model was compared to the *Baseline* model that was trained with all of the first five days for the same total number of epochs (30). Despite the same number of epochs, the *Baseline* model has been exposed to substantially more data, since 80% of the *Baseline* training data (5 days) was shown for 30 epochs as compared to 80% of the *CL_full* initial training data (4 days) for 25 epochs, plus 80% of the secondary training data (1 day) for 5 epochs. The result is that the *CL_full* model had only 70% of the data exposure as compared to the *Baseline* model. Figures 2a and 2b show that the effect the curriculum learning approach has on model performance, with a stark improvement in performance on evaluation days. In addition to the improvement in performance, the total training time for the model using curriculum learning was 25 total hours vs 28 hours for the *Baseline* method.

Model Name	**Train Days**	**Num Epochs**	**Train Days**	**Num Epochs**	**Train Time**
CL_full	[0,1,2,3]	25	4	5	24+1H
CL_less	[0,1,2,3]	20	4	5	20+1H
Baseline	[0,1,2,3,4]	30	-	-	28H

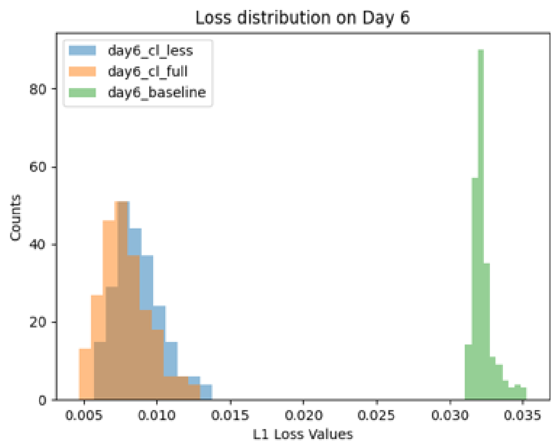
Table 1: For the curriculum learning experiment, we trained the *CL_full*, *CL_less*, and *Baseline* models using three different curricula (i.e. training sequences). The *Baseline* model was trained using the most total training data - all 5 days of data at once - and has the correspondingly longest training time. The two *CL* models were shown the first four days of data for varying numbers of epochs and then were both trained on the fifth day of data. All models were evaluated using data from days 5 and 6, and those results are shown in Figures 2a and 2b, respectively.

To continue to investigate the potential of this approach, we conduct a second experiment to compare the curriculum learning paradigm when using even less training (*CL_less* model). In figures 2a and 2b, we see that while the *CL_less* model is outperformed by *CL_full*, it still outperforms the baseline, despite having 20% fewer epochs of training. That reduction in epochs results in the model being shown only 80% the amount of data as compared to the *CL_full* model. The *CL_less* model took 21 hours to train the entire curricula, 4 hours fewer than *CL_full*.

These experiments demonstrate a successful example of using a curriculum learning approach to reduce the error when forecasting on future days. In the first experiment, this improvement in performance is shown despite equal amounts of training epochs and less training data exposure, and the second experiment shows a similar result despite even fewer training epochs and training data exposure. This approach could have potential impact for forecasting in operational environments, particularly since the resulting models are both faster to train and exhibit better performance than the *Baseline* model.



(a) Evaluation Results on Day 5 with and without performing curriculum learning during training.



(b) Evaluation Results on Day 6 with and without performing curriculum learning during training.

Figure 2: Distribution of loss values after performing curriculum learning (*CL_full* and *CL_less*), which improves performance compared to the *Baseline* approach.

Simulated Online Learning Experiment with and without GPUs

For our second experiment, we simulate an Online Learning scenario where new data can be used to refresh an older model that has been trained on data that is potentially out of distribution relative to current data. We use data from one day in September 2022 and train the model for 5 epochs to show that the Online Learning paradigm can leverage models trained on historical data and update them to provide effective results faster rather than starting from scratch.

The data from September 2022 was collected, aggregated, and processed in the same way as the September 2020 data in the first experiment. It is important to note that there were more extreme air quality events present in the September 2020 training data (see Table 2 in the Appendix), so tuning this model to a more reasonable, current state mimics a

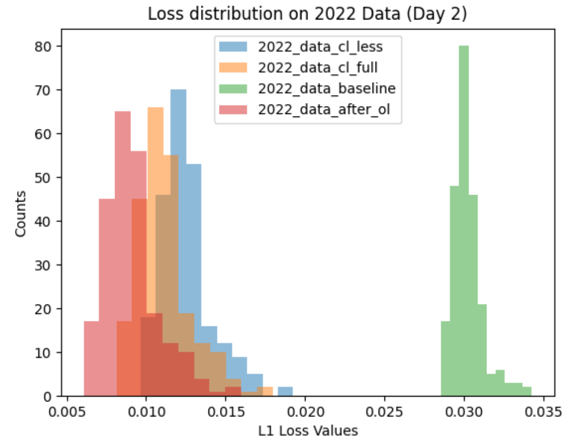


Figure 3: Distribution of loss values on Day 2 from all three models from Experiment 1. The best performing model, *CL_full* was chosen to simulate an Online Learning scenario with new data coming from 2 years after data the model was initially trained with, and shows an improvement in performance with just 5 epochs of training on data from the same month.

scenario where we may only have a small amount of data or resources available to update an older model and need to identify the fastest way to achieve the best performance.

To test the feasibility of online learning in operationally-constrained environments, we repeated this experiment without using GPU resources. Instead, an Intel Neon CPU was used. Results show that the online learning completes training in 2 hours on the CPU as compared to 1 hour on the NVIDIA H100 GPU with predictive performance maintained. Coupling the resilience method described above with the fine-tuning of models on CPU, we show, in this limited air quality study, the feasibility of using continual learning for operational environments that are resource-constrained.

Conclusions and Discussion

In this work, we found that the lightweight CL techniques hold promise for deployed systems that need to be robust to distribution shift, but may not be able to accommodate a more complex approach involving DL model updates. This exploration is an example of two techniques that have the potential to be leveraged to a.) increase the resilience of FAW systems and b.) enable online learning in resource-constrained environments. We show that the CL approaches were able to improve performance using less training data, in faster time, and when using non-GPU environments, still maintained a reasonable time performance for online learning. Future work will investigate more sophisticated curricula and other lightweight approaches, particularly tuned for distribution shifts that often occur when extreme air quality events are present. These techniques hold promise for application to other types of resource-constrained applications.

Funding Disclosure

This work has been funded by the following NOAA grant: NA22NES4050023D, SUBAWD118488-Z7642201.

References

Abel, D.; Barreto, A.; Van Roy, B.; Precup, D.; van Hasselt, H. P.; and Singh, S. 2024. A definition of continual reinforcement learning. *Advances in Neural Information Processing Systems*, 36.

Ahmad, S.; Shakeel, I.; Mehfuz, S.; and Ahmad, J. 2023. Deep learning models for cloud, edge, fog, and IoT computing paradigms: Survey, recent advances, and future directions. *Computer Science Review*, 49: 100568.

Ao, S.-I.; and Fayek, H. 2023. Continual deep learning for time series modeling. *Sensors*, 23(16): 7167.

Bahari, N. A. A. B. S.; Ahmed, A. N.; Chong, K. L.; Lai, V.; Huang, Y. F.; Koo, C. H.; Ng, J. L.; and El-Shafie, A. 2023. Predicting sea level rise using artificial intelligence: a review. *Archives of Computational Methods in Engineering*, 30(7): 4045–4062.

Baker, M. M.; New, A.; Aguilar-Simon, M.; Al-Halah, Z.; Arnold, S. M.; Ben-Iwhiwhu, E.; Brna, A. P.; Brooks, E.; Brown, R. C.; Daniels, Z.; et al. 2023. A domain-agnostic approach for characterization of lifelong learning systems. *Neural Networks*, 160: 274–296.

Ben Ayed, R.; and Hanana, M. 2021. Artificial intelligence to improve the food and agriculture sector. *Journal of Food Quality*, 2021(1): 5584754.

Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, 41–48.

Besnard, Q.; and Ragot, N. 2024. Continual Learning for Time Series Forecasting: A First Survey. *Engineering Proceedings*, 68(1): 49.

Chen, R. R.; Ribaudo, C.; Sleeman, J.; Ashcraft, C.; Kofroth, C.; Hughes, M.; Stajner, I.; Viner, K.; and Wang, K. 2024. Difference Learning for Air Quality Forecasting Transport Emulation. *arXiv preprint arXiv:2402.14806*.

Chen, Z.; and Liu, B. 2022. *Lifelong machine learning*. Springer Nature.

Csurka, G. 2017. Domain adaptation for visual applications: A comprehensive survey. *arXiv preprint arXiv:1702.05374*.

Faber, K.; Zurek, D.; Pietron, M.; Japkowicz, N.; Vergari, A.; and Corizzo, R. 2024. From MNIST to ImageNet and back: benchmarking continual curriculum learning. *Machine Learning*, 1–28.

Hadsell, R.; Rao, D.; Rusu, A. A.; and Pascanu, R. 2020. Embracing change: Continual learning in deep neural networks. *Trends in cognitive sciences*, 24(12): 1028–1040.

Hoi, S. C.; Sahoo, D.; Lu, J.; and Zhao, P. 2021. Online learning: A comprehensive survey. *Neurocomputing*, 459: 249–289.

Jaffe, D. A.; O'Neill, S. M.; Larkin, N. K.; Holder, A. L.; Peterson, D. L.; Halofsky, J. E.; and Rappold, A. G. 2020. Wildfire and prescribed burning impacts on air quality in the United States. *Journal of the Air & Waste Management Association*, 70(6): 583–615.

Koenecke, A.; and Gajewar, A. 2020. Curriculum learning in deep neural networks for financial forecasting. In *Mining Data for Financial Applications: 4th ECML PKDD Workshop, MIDAS 2019, Würzburg, Germany, September 16, 2019, Revised Selected Papers 4*, 16–31. Springer.

Leal Filho, W.; Wall, T.; Mucova, S. A. R.; Nagy, G. J.; Balogun, A.-L.; Luetz, J. M.; Ng, A. W.; Kovaleva, M.; Azam, F. M. S.; Alves, F.; et al. 2022. Deploying artificial intelligence for climate change adaptation. *Technological Forecasting and Social Change*, 180: 121662.

Menghani, G. 2023. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, 55(12): 1–37.

Misra, A. K. 2014. Climate change and challenges of water and food security. *International Journal of Sustainable Built Environment*, 3(1): 153–165.

Morton, J. F. 2007. The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the national academy of sciences*, 104(50): 19680–19685.

Parisi, G. I.; Kemker, R.; Part, J. L.; Kanan, C.; and Wermter, S. 2019. Continual lifelong learning with neural networks: A review. *Neural networks*, 113: 54–71.

Pham, Q.; Liu, C.; Sahoo, D.; and Hoi, S. C. 2022. Learning fast and slow for online time series forecasting. *arXiv preprint arXiv:2202.11672*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Schillaci, G.; Schmidt, U.; and Miranda, L. 2021. Prediction Error-Driven Memory Consolidation for Continual Learning: On the Case of Adaptive Greenhouse Models. *KI-Künstliche Intelligenz*, 35(1): 71–80.

Stajner, I. 2022. Development of Next-Generation Air Quality Predictions for the United States. In *EGU General Assembly Conference Abstracts*, EGU22–10491.

Stojnic, R.; Taylor, R.; Kardas, M.; Kerkez, V.; and Viaud, L. 2022. Papers with Code-The latest in Machine Learning. *URL: https://paperswithcode. com*.

Wang, L.; Zhang, X.; Su, H.; and Zhu, J. 2024. A comprehensive survey of continual learning: theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wiebe, K.; Robinson, S.; and Cattaneo, A. 2019. Climate change, agriculture and food security: impacts and the potential for adaptation and mitigation. *Sustainable food and agriculture*, 55–74.