

---

# Short-range forecasts of global precipitation using deep learning-augmented numerical weather prediction

---

**Manmeet Singh**

Jackson School of Geosciences  
The University of Texas at Austin,  
Austin, TX 78712  
manmeet.singh@utexas.edu

**Vaisakh SB**

Indian Institute of Tropical Meteorology  
Ministry of Earth Sciences  
Pune, India 411008  
vaisakh.sb@tropmet.res.in

**Nachiketa Acharya**

CIRES, University of Colorado Boulder  
NOAA/Physical Sciences Laboratory  
Boulder, CO 80309  
dr.nachiketaacharya@gmail.com

**Aditya Grover**

Department of Computer Science  
University of California, Los Angeles  
Los Angeles, CA 90095  
adityag@cs.ucla.edu

**Suryachandra A. Rao**

Indian Institute of Tropical Meteorology  
Ministry of Earth Sciences  
Pune, India 411008  
surya@tropmet.res.in

**Bipin Kumar**

Indian Institute of Tropical Meteorology  
Ministry of Earth Sciences  
Pune, India 411008  
bipink@tropmet.res.in

**Zong-Liang Yang**

Jackson School of Geosciences  
The University of Texas at Austin,  
Austin, TX 78712  
liang@jsg.utexas.edu

**Dev Niyogi**

Jackson School of Geosciences  
The University of Texas at Austin,  
Austin, TX 78712  
dev.niyogi@jsg.utexas.edu

## Abstract

Precipitation drives the hydroclimate of Earth and its spatiotemporal changes on a day to day basis have one of the most notable socioeconomic impacts. The success of numerical weather prediction (NWP) is measured by the improvement of forecasts for various physical fields such as temperature and pressure. Large biases however exist in the precipitation predictions. Pure deep learning based approaches lack the advancements achieved by NWP in the past two to three decades. Hybrid methodology using NWP outputs as inputs to the deep learning based refinement tool offer an attractive means taking advantage of both NWP and state of the art deep learning algorithms. Augmenting the output from a well-known NWP model: Coupled Forecast System ver.2 (CFSv2) with deep learning for the first time, we demonstrate a hybrid model capability (**DeepNWP**) which shows substantial skill improvements for short-range global precipitation at 1-, 2- and 3-days lead time. To achieve this hybridization, we address the sphericity of the global data by using modified DLWP-CS architecture which transforms all the fields to cubed-sphere projection. The dynamical model outputs corresponding to precipitation and surface temperature are ingested to a UNET for predicting the target ground truth precipitation. While the dynamical model CFSv2 shows

a bias in the range of +5 to +7 mm/day over land, the multivariate deep learning model reduces it to -1 to +1 mm/day over global land areas. We validate the results by taking examples from Hurricane Katrina in 2005, Hurricane Ivan in 2004, Central European floods in 2010, China floods in 2010, India floods in 2005 and the Myanmar cyclone Nargis in 2008.

# 1 Introduction

Precipitation forms an essential component of the hydrological cycle for the planet. The precipitation falling on land and oceans provides sustenance to the living beings and its accurate predictions can lead to tremendous societal benefits. At short-range time scales, i.e. of the order one to three days, the information of rain occurrence is helpful for various stakeholders. For example, ahead in time knowledge of precipitation can help better mitigate or manage the impacts of hurricanes, cyclones and floods. At present, most of the weather forecasting in the world is performed by the models which are known as dynamical models or numerical weather prediction (NWP) models (2). Past few decades have seen improvement in the skill of numerical weather predictions, particularly that of temperature and winds. However, when it comes to physical fields such as precipitation and soil moisture, there are large biases in the outputs of these models (27). Deep learning can be effectively used as a tool to correct the biases in NWP models (22). The ability of deep learning to unwrap nonlinear patterns in the data coupled with its success in the spatial tasks within the computer vision community makes it specially attractive for this task (15). There have been attempts to perform data-driven weather forecasting by various researchers in the past (13; 23). They were however limited by (i) performing training on linear fields relative to precipitation such as geopotential height at 500 hPa; thus not effectively addressing the physical field of direct interest to the humans (3; 4; 5; 20; 25; 26; 32; 33), (ii) their ability to address sphericity of the global data over Earth (1; 3; 5; 10; 17; 20; 21; 21; 25; 26; 29; 31; 36; 18), (iii) performing their study only over a limited region; thus neglecting the global teleconnections (1; 3; 21; 29; 18), (iv) their framework not utilizing and building upwards on the advancements in NWP; and rather trying to achieve the results similar to NWP and not utilizing the advancements in NWP & deep learning to develop hybrid forecasting framework to enhance the skill of the existing systems (1; 3; 4; 5; 17; 20; 29; 32; 33; 36).

## 1.1 Our contributions

Rather than using deep learning directly to forecast precipitation, we use a hybrid approach combining it with the traditional NWP. In particular, we consider CFSv2 which is a model used for operational weather prediction in the United States, India and other countries. We propose to correct its biases by training a deep neural network that projects its outputs to ERA5 reanalysis, which is a merged product of ground and space-based observations and data-assimilation available from European Centre for Medium-Range Weather Forecasts is used as a target. ERA5 reanalysis precipitation has shown high fidelity in representing precipitation in the recent times (8). Thus, we use the outputs of precipitation and temperature from CFSv2 model and map them to the ERA5 reanalysis precipitation. The architecture of deep learning based computer vision algorithm used in modified DLWP-CS (30).

Our NWP augmented deep learning product shows substantial improvements in the representation of precipitation. On an average, we are able to reduce the bias in precipitation from beyond +10 mm/day to within  $\pm 1$  mm/day. The results are consistent for different lead days in the short range time scale (+1, +2 and +3 days). When segregated based on the different seasons, the global mean bias values show improvement upwards of 10x by introducing deep learning in the numerical weather prediction system. We validate our results with specific use cases comprising of heavy to extreme precipitation from the test data. The use cases corroborate the findings from the global mean bias and show even more substantial improvements, the maximum being for India flood 2005 wherein the bias in NWP decreased from +17.321 mm/day to +0.003 mm/day for day 1 lead forecast when deep learning is added to the system. Our hybrid model can be used for mitigating the impacts of various natural disasters triggered by precipitation and effective management of water resources.

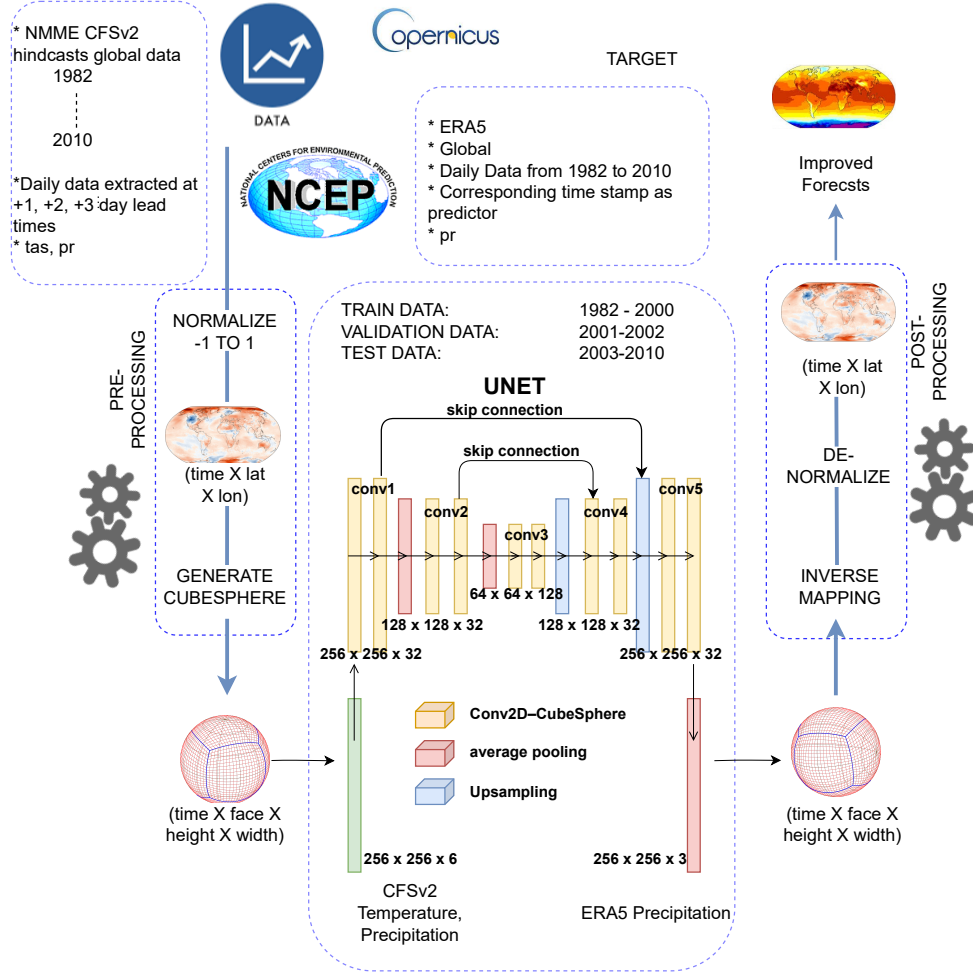


Figure 1: Schematic of the deep learning-augmented numerical weather prediction.

## 2 Data and methodology

The input dataset comprises of precursors, viz, CFSv2 generated precipitation and surface air temperature outputs. Precipitation and temperature outputs of CFSv2 are downloaded from North American Multi-Model Ensemble (NMME) for developing the input training dataset. ERA5 reanalysis (11) is used as the ground truth representing the real-world precipitation. We obtain the precursors from the CFSv2 model outputs and the target extracted from ERA5 reanalysis. The CFSv2 output is available as 6-hourly average for a lead time of upto +10 months while the ERA5 reanalysis is available as an hourly product. Daily aggregates are first generated for the input (CFSv2) and target (ERA5) by summing the sub-daily datasets. Then, the time slices which represent 1, 2 and 3 days lead in CFSv2 are extracted. For the labels from ERA5, the data is extracted such that it corresponds to the forecasted 1, 2 and 3 days from CFSv2 in reality. Data from 1982 to 2000 is used for training, 2000-2001 for validation and 2003 to 2010 for testing. The maximum and minimum values of the training data fields are then computed for normalizing the data. Normalization scales the different variables to the range -1 to +1. We call our model as **DeepNWP** to signify the hybrid nature of deep learning augmented numerical weather prediction. The methodology used is shown as a schematic in figure 1

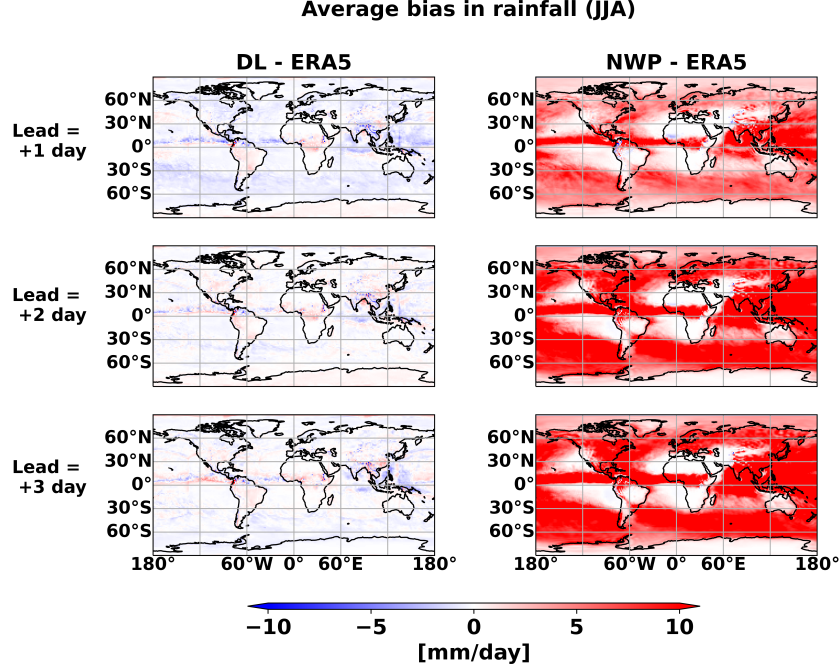


Figure 2: Average bias in precipitation for June to August season for the deep learning augmented NWP model (DL) and NWP model relative to ERA5 reanalysis. The rows represent the bias at lead time = 1, 2 and 3 days.

Season	Lead = 1 day (mm/day)		Lead = 2 day (mm/day)		Lead = 3 day (mm/day)	
	DL - ERA5	CFSv2 - ERA5	DL - ERA5	CFSv2 - ERA5	DL - ERA5	CFSv2 - ERA5
DJF	<b>-0.3</b>	3.827	<b>0.022</b>	7.66	<b>-0.158</b>	7.657
MAM	<b>-0.282</b>	3.834	<b>0.032</b>	7.811	<b>-0.11</b>	7.89
JJA	<b>-0.334</b>	3.97	<b>-0.02</b>	8.102	<b>-0.115</b>	8.239
SON	<b>-0.299</b>	3.954	<b>0</b>	7.95	<b>-0.148</b>	7.972

Table 1: Performance of the deep learning augmented numerical weather prediction system DeepNWP (bold) versus CFSv2 alone. The table shows global average bias/error in simulating precipitation by the hybrid deep learning and CFSv2 system versus CFSv2 alone. DJF (December to February), MAM (March to May), JJA (June to August) and SON (September to November) represent the different months of an year. The performance is shown for the entire test period from the year 2003 to 2010.

### 3 Results

Figure 2 shows the performance measured by average bias or error in deep learning augmented NWP model (DL) and NWP model alone relative to ERA5 reanalysis during June to August season. The first column corresponds to DL bias and second column represents the NWP model bias. The average bias or error is computed as a mean of the difference between the model (DL or NWP) and ERA5 reanalysis for all the samples in the test dataset. We also compute the plots similar to figure 2 for other seasons, viz, December to February, March to May and September to November (figures not shown). From the average seasonal bias figures, it can be noted that deep learning augmented NWP model substantially reduces the bias in precipitation relative to the NWP (CFSv2) model. While the mean bias over important land regions lies within the range +5 to +7 mm/day in NWP (CFSv2) (similar to the results of (19)), it falls to within the range -1 to +1 mm/day using deep learning. Thus, in a mean sense, deep learning augmented CFSv2 model improves the performance of precipitation forecasts by upto 4-5x. We can note the enhanced performance until the computed +3 days lead time for deep learning augmented NWP model. Table 1 shows the bias (model - ground truth) in global mean precipitation at different lead times (1, 2 and 3 days). Substantially enhanced skill of global precipitation forecasts upwards of 10x is noted for the all the seasons.

## 4 Conclusions

In this study, we develop a hybrid deep learning augmented numerical weather prediction (DeepNWP) model for generating global precipitation forecasts at short-range (1, 2 and 3 days) time scales. While previous studies have noted the possibility of a hybrid deep learning augmented NWP model and also suggesting an end-to-end deep learning based system (28; 32), ours is a first attempt in actually developing such a system. All the existing implementations using deep learning for weather prediction have attempted for relatively simplistic fields such as geopotential height, while we actually attempt to improve global precipitation forecasts. We ensure that the sphericity of the global data over Earth is considered and find vast improvements in deep learning augmented NWP forecasts relative to the NWP alone. In future, we would use more precursors of precipitation from the NWP model for generating deep learning augmented weather predictions. Further, hyperparameter tuning and neural architecture search would be performed for performance enhancement.

## References

- [1] Troy Arcomano, Istvan Szunyogh, Jaideep Pathak, Alexander Wikner, Brian R Hunt, and Edward Ott. A machine learning-based global atmospheric forecast model. *Geophysical Research Letters*, 47(9):e2020GL087776, 2020.
- [2] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.
- [3] Alex Bihlo. A generative adversarial network approach to (ensemble) weather prediction. *Neural Networks*, 139:1–16, 2021.
- [4] Alex Bihlo and Roman O Popovych. Physics-informed neural networks for the shallow-water equations on the sphere. *Journal of Computational Physics*, page 111024, 2022.
- [5] Ashesh Chattopadhyay, Pedram Hassanzadeh, and Devika Subramanian. Data-driven predictions of a multiscale lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. *Nonlinear Processes in Geophysics*, 27(3):373–389, 2020.
- [6] SA Clough, MW Shephard, EJ Mlawer, JS Delamere, MJ Iacono, K Cady-Pereira, S Boukabara, and PD Brown. Atmospheric radiative transfer modeling: A summary of the aer codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 91(2):233–244, 2005.
- [7] MB Ek, KE Mitchell, Ying Lin, Eric Rogers, Pablo Grunmann, Victor Koren, George Gayno, and JD Tarpley. Implementation of noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research: Atmospheres*, 108(D22), 2003.
- [8] Stephanie Gleixner, Teferi Demissie, and Gulilat Tefera Diro. Did era5 improve temperature and precipitation reanalysis over east africa? *Atmosphere*, 11(9):996, 2020.
- [9] Stephen M Griffies, Matthew J Harrison, Ronald C Pacanowski, and Anthony Rosati. A technical guide to mom4. *GFDL Ocean Group Tech. Rep.*, 5:342, 2004.
- [10] Yilun Han, Guang J Zhang, Xiaomeng Huang, and Yong Wang. A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002076, 2020.
- [11] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.
- [12] Michael J Iacono, Eli J Mlawer, Shepard A Clough, and Jean-Jacques Morcrette. Impact of an improved longwave radiation model, rrtm, on the energy budget and thermodynamic properties of the ncar community climate model, ccm3. *Journal of Geophysical Research: Atmospheres*, 105(D11):14873–14890, 2000.
- [13] K Kashinath, M Mustafa, A Albert, JL Wu, C Jiang, S Esmailzadeh, K Azizzadenesheli, R Wang, A Chattopadhyay, A Singh, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.
- [14] Young-Joon Kim and Akio Arakawa. Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *Journal of Atmospheric Sciences*, 52(11):1875–1902, 1995.
- [15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [16] François Lott and Martin J Miller. A new subgrid-scale orographic drag parameterization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, 123(537):101–127, 1997.
- [17] Griffin Mooers, Michael Pritchard, Tom Beucler, Jordan Ott, Galen Yacalis, Pierre Baldi, and Pierre Gentile. Assessing the potential of deep learning for emulating cloud superparameterization in climate models with real-geography boundary conditions. *Journal of Advances in Modeling Earth Systems*, 13(5):e2020MS002385, 2021.
- [18] Soukayna Mouatadid, Paulo Orenstein, Genevieve Flaspohler, Miruna Oprescu, Judah Cohen, Franklyn Wang, Sean Knight, Maria Geogdzhayeva, Sam Levang, Ernest Fraenkel, et al. Learned benchmarks for subseasonal forecasting. *arXiv preprint arXiv:2109.10399*, 2021.
- [19] P Mukhopadhyay, VS Prasad, R Krishna, Medha Deshpande, Malay Ganai, Snehlata Tirkey, Sahadat Sarkar, Tanmoy Goswami, CJ Johnny, Kumar Roy, et al. Performance of a very high-resolution global

- forecast system model (gfs t1534) at 12.5 km over the indian region during the 2016–2017 monsoon seasons. *Journal of Earth System Science*, 128(6):1–18, 2019.
- [20] Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020.
  - [21] Stephan Rasp and Sebastian Lerch. Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146(11):3885–3900, 2018.
  - [22] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.
  - [23] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.
  - [24] Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui-ya Chuang, Mark Iredell, et al. The ncep climate forecast system version 2. *Journal of climate*, 27(6):2185–2208, 2014.
  - [25] Sebastian Scher. Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters*, 45(22):12–616, 2018.
  - [26] Sebastian Scher and Gabriele Messori. Weather and climate forecasting with neural networks: using general circulation models (gcms) with different complexity as a study ground. *Geoscientific Model Development*, 12(7):2797–2809, 2019.
  - [27] Gavin Schmidt. The real holes in climate science. *Nature*, 463:21, 2010.
  - [28] MG Schultz, Clara Betancourt, Bing Gong, Felix Kleinert, Michael Langguth, LH Leufen, Amirpasha Mozaffari, and Scarlet Stadler. Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, 379(2194):20200097, 2021.
  - [29] Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in neural information processing systems*, 30, 2017.
  - [30] Manmeet Singh, Bipin Kumar, Suryachandra Rao, Sukhpal Singh Gill, Rajib Chattopadhyay, Ravi S Nanjundiah, and Dev Niyogi. Deep learning for improved global precipitation in numerical weather prediction systems. *arXiv preprint arXiv:2106.12045*, 2021.
  - [31] Rui Wang, Karthik Kashinath, Mustafa Mustafa, Adrian Albert, and Rose Yu. Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1457–1466, 2020.
  - [32] Jonathan A Weyn, Dale R Durran, and Rich Caruana. Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11(8):2680–2693, 2019.
  - [33] Jonathan A Weyn, Dale R Durran, and Rich Caruana. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002109, 2020.
  - [34] Michael Winton. A reformulated three-layer sea ice model. *Journal of atmospheric and oceanic technology*, 17(4):525–531, 2000.
  - [35] Xingren Wu, Ian Simmonds, and WF Budd. Modeling of antarctic sea ice in a general circulation model. *Journal of Climate*, 10(4):593–609, 1997.
  - [36] Janni Yuval and Paul A O’Gorman. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, 11(1):1–10, 2020.

## Appendix A Training details

The training is performed on a single NVIDIA A100 GPU with 40 GB GPU RAM. Since the size of data is huge for I/O, loading on GPU and training, efficient use of TensorFlow dataloaders is made. The total number of training samples is 5528 corresponding to the data from 1982 to 2000 with a batch size of 8. The batch size is chosen to ensure that the heavy data can be trained on the A100 GPU by iterating on different choices. We use the adam optimizer, learning rate of 0.0001 and mean squared error as loss function. Since I/O was a big overhead for the training, it is performed in loop by training each year in one go and then improving the model for years that come after. We call it incremental training and to prevent overfitting, early stopping is used.

Incorporating the effects of sphericity while also using a computer vision based deep learning model, the latitude-longitude dataset is transformed to the cubed sphere mapping. We use the implementation of tempest remap to perform the cubed sphere transformation. Cubed sphere is typically a cubical sphere with 6 faces. Two of these faces represent the polar regions while four faces are over the tropics. The resolution of each face of the cubed sphere that is used in this study is 256 x 256. The choice of the cubed sphere resolution is determined by transforming from a regular latitude-longitude grid to cubed sphere and then back to the regular latitude longitude grid. Errors induced by the cubed

sphere by a forward and backward pass are computed and the optimal cubed resolution of 256 x 256 is selected.

We use modified DLWP-CS (30) which is evolved from DLWP-CS (33) for training the deep learning model. Modified DLWP-CS use a 2D UNET to perform an image to image regression for the cubed sphere data to map from one or multiple variables to uni or multivariate target. The core of modified DLWP-CS uses two 2D UNETs, one for the tropics and other for the poles. They are coupled to each other in such a way that the data is exchanged while ensuring the same input and output matrix size during convolution. Typically, in computer vision applications, a padding of zeros is used. However, a global spherical data doesn't have any real boundaries, so neighbouring edge of a cubed sphere face is used for padding. This data exchange process ensures that the important global linkages such as those to and from El Nino Southern Oscillation, North Atlantic Oscillation and Pacific Decadal Oscillation are considered by the data-driven approach. The different input variables from CFSv2 model; viz, precipitation and temperature are fed as channels to the UNET from modified DLWP-CS. The target is a single channel ERA5 precipitation in cubed sphere projection.

## **Appendix B Preliminaries**

Decades of scientific research into improving the global precipitation forecasts has led to the state of the art NWP systems which are behind the operational weather prediction worldwide. Although, NWP has been a success, problems remain in fields such as precipitation and soil moisture showing high bias in their representation in the NWP models. The NWP systems are basically solutions of governing equations of the atmosphere ocean and land which are coupled to each other. These partial differential equations are solved by assuming Earth to be a sphere and initializing from appropriate initial and boundary conditions. The equations are solved for prognostic fields such as pressure, temperature, winds and tracers such as water vapor and ozone. Precipitation and soil moisture are diagnostic fields which are derived by empirical relationships also known as parameterization schemes taking in the prognostic fields as input. These parameterization schemes are challenging to build and improve in the NWP systems and the problems with diagnostic fields are mostly attributed to them. CFSv2 is a model which is used operationally for weather forecasting by different nodal centers around the world.

Recently, studies (13; 23) have shown the capability of deep learning to emulate weather systems purely from a data-driven approach. However, these methods often lack the capability to deliver a global product which can be directly used for societal applications. More on the limitations of these studies has been discussed in section 1.1. Applications (32; 33) have attempted to solve various challenges in pure data-driven models such as addressing the sphericity of Earth (33) and using global datasets. However, they (32; 33) lack the strength of a unified deep learning NWP system. We use ERA5 reanalysis global precipitation as the ground truth target. ERA5 reanalysis is a merged product of satellite, gauge-based ground observations, ocean buoys and other instrumental data combined with model outputs to generate a global product. It has shown fidelity to represent precipitation similar to the observations.

## **Appendix C Climate Forecast system version 2 (CFSv2)**

CFSv2 is a spectral model in which atmosphere, ocean and land components are coupled to each other. Within the CFSv2, the atmospheric model is known as the Global Forecast System (GFS). The atmospheric component of CFSv2, i.e. GFS has a spectral resolution has a spectral resolution of T126 with 64 vertical levels (24). Ocean model of CFSv2 is the Modular Ocean Model developed by the Geophysical Fluid Dynamics Laboratory (GFDL) (9). Arakawa-Schubert scheme is used for convective parameterization with momentum mixing. CFSv2 incorporates the effects such as orographic gravity wave drag and mountain blockage (14; 16). Rapid radiative transfer model is used for the radiation computations in the atmosphere (12; 6). A four-layer land-surface model known as the NOAA LSM (7) and a dynamical two-layer sea ice model (35; 34) are coupled to the atmosphere and ocean components of CFSv2. CFSv2 hindcast simulations from 1982 to 2010 are downloaded. The hindcast model simulations are initialized at an interval of 5 days for the years 1982-2010. A total number of 4 ensemble members are generated for each initialization day of the model. We use all the available ensembles for training.

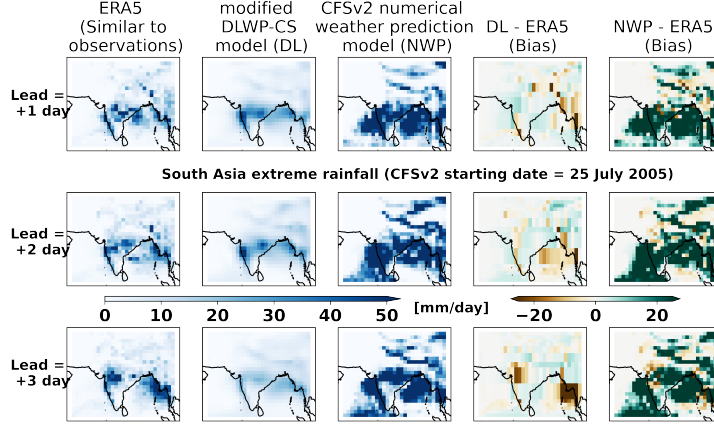


Figure 3: Absolute values corresponding to ERA5, deep learning augmented CFSv2 (DL) and CFSv2 are shown in the first three columns. Last two columns show the bias as difference between deep learning augmented CFSv2 and CFSv2 alone. The rows correspond to the different lead times, viz, 1, 2 and 3 days. The figure shows example for South Asia extreme precipitation in 2005.

## Appendix D ERA5 reanalysis

ERA5 dataset is provided by the European Center for Medium-Range Weather Forecasts (ECMWF) and has been available since 2019. It consists of meteorological fields at hourly temporal intervals and a global spatial resolution of 0.25 degrees. The data is available from the year 1979 which was one of the reason for selecting ERA5 reanalysis as the ground truth. Other global datasets such as GPCC and GPCP are either available since the satellite era (1996 onwards) precipitation products became available or are only over land. Recent studies have shown that ERA5 reanalysis precipitation provides a reasonable representation of the precipitation (8).

## Appendix E Case studies

We validate the skill of **DeepNWP** for specific examples viz, (i) Hurricane Katrina, (ii) Hurricane Ivan, (iii) Cyclone Nargis, (iv) Europe flood 2010, (v) China flood 2010 and (vi) India flood 2005. Spatial maps of absolute values corresponding to ERA5 (similar to observations), deep learning augmented CFSv2 and CFSv2 are shown at the lead time = 1, 2 and 3 days. The lead times are with reference to the starting date of CFSv2 mode, i.e. since the time CFSv2 gets initialized. In addition, the figures show difference in model predicted precipitation relative to ERA5 for the three considered lead times. Considering the individual hurricane, cyclone and flooding events, it can be seen that CFSv2 has a much larger bias in the tropics relative to mid latitudes (figures 4, 5, 6, 3, 7). CFSv2 has a wet bias for the different events, even exceeding +20 mm/day over different regions across the selected cases. DeepNWP reduces the bias to within the range -1 to +1 mm/day except for cyclone Nargis. Cyclone Nargis shows humongous bias in CFSv2 from +20 to +47 mm/day while the deep learning model reduces these errors from -5.37 to +2.34 mm/day for the three lead times considered. A summary of these statistics is provided in the table 2.

Following are some case studies from the test data period corresponding to the years 2003-2010 showing the superior performance of deep learning augmented NWP:

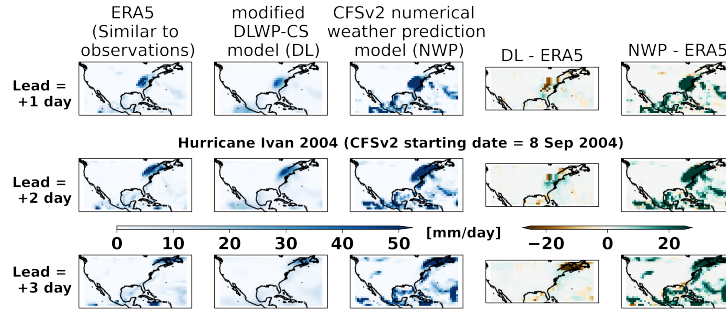


Figure 4: Same as figure 3 but for Hurricane Ivan in 2004.

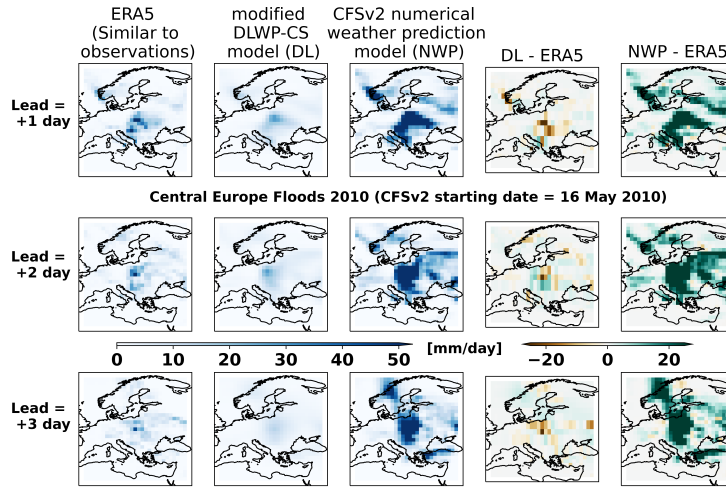


Figure 5: Same as figure 3 but for Europe flood in 2010.

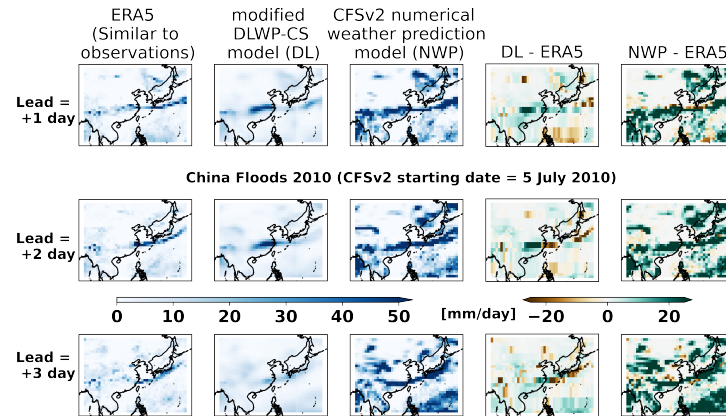


Figure 6: Same as figure 3 but for China floods in 2010.

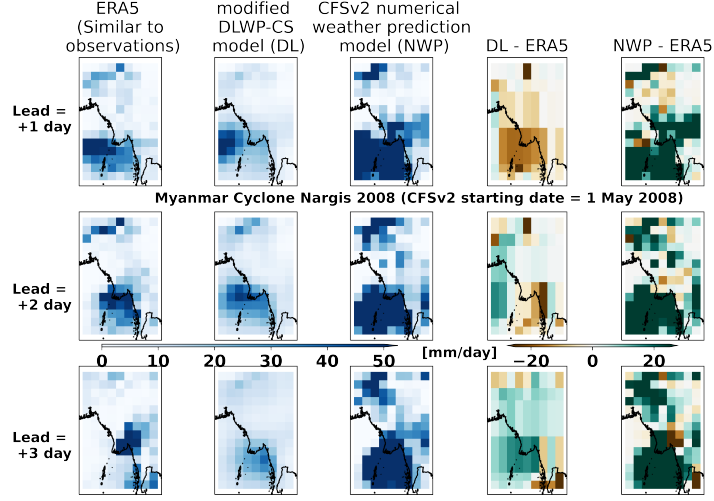


Figure 7: Same as figure 3 but for Myanmar cyclone Nargis that occurred in 2008

Events	Lead = 1 day (mm/day)		Lead = 2 day (mm/day)		Lead = 3 day (mm/day)	
	DL - ERA5	CFSv2 - ERA5	DL - ERA5	CFSv2 - ERA5	DL - ERA5	CFSv2 - ERA5
Hurricane Katrina	<b>-0.345</b>	8.839	<b>0.453</b>	12.18	<b>-0.811</b>	10.227
Hurricane Ivan	<b>-0.22</b>	8.466	<b>-0.036</b>	13.48	<b>-1.485</b>	13.135
Cyclone Nargis	<b>-5.37</b>	21.151	<b>-1.245</b>	43.845	<b>2.338</b>	47.233
Europe Floods	<b>-0.2</b>	6.654	<b>-0.015</b>	8.134	<b>0.12</b>	6.94
China Floods	<b>-0.17</b>	11.233	<b>0.465</b>	18.903	<b>-0.48</b>	16.877
India flood	<b>0.003</b>	17.321	<b>0.139</b>	25.297	<b>-0.749</b>	20.259

Table 2: Performance of the deep learning augmented numerical weather prediction system CFSv2 versus CFSv2 alone. The table shows regional bias/error in simulating various extreme precipitation events by the hybrid deep learning and CFSv2 system versus CFSv2 alone. The events occurred as (i) Hurricane Katrina in 2005, (ii) Hurricane Ivan in 2004, (iii) Cyclone Nargis in 2008, (iv) Europe floods in 2010, (v) China flood in 2005 and (vi) India flood in 2005