Comparing skill of historical rainfall data based monsoon rainfall prediction in India with NWP forecasts

Aastha Jain,[a] Apoorva Narula,[b] Jatin Batra,[c] M Rajeevan,[d] Sandeep Juneja,[a]

[a] *Ashoka University*

[b] *Indian Institute of Technology Delhi*

[c] *Tata Institute of Fundamental Research*

[d] *Atria University*

Corresponding author: Aastha Jain, aasthajn25@gmail.com

ABSTRACT: We consider the problem of forecasting rainfall across India during the four monsoon months, one day as well as three days in advance. We train neural networks using historical daily gridded precipitation data for India obtained from IMD for the time period 1901 – 2023, at spatial resolutions of $0.25^\circ \times 0.25^\circ$ as well as $1^\circ \times 1^\circ$. Additional weather variables from ECMWF (European Centre for Medium-Range Weather Forecasts) reanalysis data, including cloud cover, humidity, temperature, soil moisture, vorticity, and wind speed, are also incorporated into the models. The forecasts from these models at $0.25^\circ \times 0.25^\circ$ are compared with the numerical weather prediction (NWP) forecasts obtained from ECMWF HRES (High Resolution Ensemble System), and at $1^\circ \times 1^\circ$ with those from NCEP (National Centre for Environmental Prediction) available for the period 2011-2023. We conduct a detailed country wide analysis, as well as, separately analyze some of the most populated cities in India. Our conclusion is that forecasts obtained by applying transformer-based and other deep learning methods to historical rainfall data and other weather variables are more accurate compared to NWP forecasts as well as predictions based on persistence. Specifically, our forecasts are better than HRES forecasts and significantly outperform NCEP forecasts. On average, compared to our predictions, forecasts from HRES and NCEP model have about 22% and 43% higher error, respectively, for a single day prediction, and over 27% and 66% higher error respectively, for a three day prediction. Similarly, persistence estimates report a 40% higher error in a single day forecast, and over 69% error in a three day forecast. We observe that deep learning based forecasts also exhibit a better balance of tracking heavy rainfall and giving false positives compared to other methods. We further observe that data up to 20 days in the past is useful in reducing errors of one and three day forecasts, when a transformer based learning architecture, and to a lesser extent when an LSTM is used. Reassuringly, this information in the past data is more pronounced for landlocked regions compared to coastal ones. A key conclusion suggested by our analysis is that NWP forecasts can be substantially improved through more and diverse data relevant to monsoon prediction combined with carefully selected neural network architecture.

1. Introduction

Accurate rainfall prediction in India during monsoons is crucial for a variety of reasons - for agriculture planning, disaster management, day to day transportation planning and so on. Anecdotally, it is well known that international numerical weather predictors (NWP) do not perform well in rainfall prediction for India (see Rajeevan (23-12-2023)). It is also conjectured that during monsoons, rainfall data across India has spatial-temporal memory so that information on rainfall early on in neighbouring parts may be useful for future rainfall prediction (see Goswami and Xavier (2003)). Moreover, rainfall has been shown to be also affected by a variety of other atmospheric, land and ocean variables (Prasad and Singh 1988), such as temperature, wind, soil moisture, etc.

In this paper, we consider daily gridded precipitation data from India Meteorological Department (IMD) (Pai et al. 2014) available from 1901 - 2023, at a spatial resolution of $0.25° \times 0.25°$, as well as $1° \times 1°$ (each degree roughly corresponds to 111 km). We use this to predict rainfall for all of India, one day as well as three days into the future. We also use daily atmospheric and land data as additional covariates in an attempt to improve our forecasts. We compare our performance with operational NWP forecasts including HRES-IFS (High Resolution Integrated Forecast System) from ECMWF (European Centre for Medium-Range Weather Forecasts) (ECMWF 2021), and those from NCEP (National Centre for Environmental Prediction), see (Copernicus Climate Change Service 2018). HRES is widely regarded as the top operational weather forecasting system in the world (Lam et al. 2023). NCEP is used as it forms the base model for IMD Global Forecasting System (see (IMD 19-04-2017))

We compare and contrast the deep-learning based forecasts generated using historical rainfall data from IMD (referred to as DL-HD forecasts) and the forecasts generated using IMD rainfall data and additional covariates (called DL-HD+Covariates), with the NCEP-NWP and HRES-NWP forecasts. In an attempt to arrive at improved forecasts, we also combine the NWP forecasts with DL-HD+Covariates using a simple neural network, to generate ensemble forecasts. Much of the rainfall in India occurs during the four monsoon months June, July, August and September (JJAS), and we focus our forecasts on these. The results for experiments at $1° \times 1°$ resolution are given in the Appendix. All our experiments involving deep learning are transformer based, except in Figure 9, where we compare the performance of transformer based DL-HD forecasts with LSTM-based forecasts.

Our key observations based on forecasts for all of India as well as for 20 most populated cities, are:

1. Both HRES-NWP and NCEP-NWP forecasts exhibit higher errors compared to those based on historical rainfall data (DL-HD forecasts) and those incorporating additional covariates (DL-HD + Covariates). Across all of India, HRES-NWP forecasts show 22% and 27.2% higher errors for 1-day and 3-day forecasts, respectively, compared to DL-HD + Covariates. Similarly, NCEP-NWP forecasts have 43.2% and 65.9% higher errors for 1-day and 3-day predictions, respectively. The results for 1-day and 3-day predictions over India are detailed in Tables 2 and 3 respectively for 0.25° resolution, and in Tables A1 and A2 respectively for 1° resolution.

2. Forecasts based on persistence (where the rainfall observed on a given day is the forecast for rainfall the next day, as well as the next three days) perform poorly for all over India. Persistence estimates for all of India on average, compared to DL-HD + Covariates, report 39.3% higher error in a single day forecast, and 69.5% error in a three day forecast.

3. We also test to see if DL-HD+Covariates forecasts can be combined with NWP forecasts to arrive at further improvements in our predictions. While the overall improvement in terms of error is only marginal, in some cities we see that the ensemble performs better.

4. We compare the ability of different forecasts to warn of heavy rainfall (if the actual rainfall in a day was above 30 mm and the forecast was higher than 10 mm) as well as to raise false alarms (if the forecast was above 30 mm and the actual rainfall was less than 10 mm). For 3-day forecasts, the lower and upper thresholds are 20 mm and 60 mm, respectively. These thresholds could not be made higher as the amount of data available was insufficient for higher values in most regions. We find that while HRES-NWP forecasts show a balanced performance with 81.9% accuracy for heavy rainfall prediction and a 41.2% false alarm rate, NCEP-NWP forecasts tend to overestimate 1-day rainfall. This results in a higher accuracy for heavy rainfall prediction at 83.9% but a poorer performance on the false alarm metric at 43.2%. In contrast, DL-HD + Covariates reports the smallest false alarm rate of 27.1% while still maintaining a high rate of heavy rainfall prediction at 82.2%. They also show a high correlation of 0.82 with the IMD ground truth data. For 3-day forecasts, the ensemble of

DL-HD + Covariates and NWP outperforms the other models both in heavy rainfall prediction accuracy (74.8%) and in minimizing false alarms, with a 37.4% miss rate. These results are detailed in Table 6 for experiments at 0.25° resolution, and in Table A3 for 1° resolution.

5. We also analyze the results for the 20 most populated cities in India. We find that DL-HD+Covariates performs slightly better in these cities compared to its performance across the entire country. On average, in these cities, HRES and NCEP report 27.8% and 46% higher errors, respectively, than forecasts from DL-HD + Covariates for a single-day prediction. For a 3-day prediction, HRES and NCEP report 28.4% and 47.2% higher errors, respectively. The results for these select cities are given in Tables 4 and 5 for 1-day and 3-day predictions respectively, for 0.25° resolution, and Tables A4 and A5 for 1° resolution.

6. Rainfall in India, as recorded by IMD, may exhibit long-term memory. We observe that forecasts improve as the historical rainfall data context increases from 3 to 20 days, particularly when using the transformer-based learning architecture. The improvement is less pronounced when Long-Short Term Memory (LSTM) networks are used. Additionally, the long-term memory effect is evident for key cities in India. Specifically, forecasts for landlocked cities and region improve as the past data context increases to about 20 days, whereas there is no significant improvement for coastal regions. This is illustrated in Figures 9 and 10.

Literature Review

Several attempts have been made to predict rainfall using machine learning (ML) techniques. For long-range forecasting of monsoon rainfall in India, Rajeevan et al. (Rajeevan et al. 2000, 2007) employed a host of methods such as multivariate principal component regression, simple neural networks, linear discriminant analysis, ensemble multiple linear regression and projection pursuit regression. They used multimodal data such as air temperature, sea surface temperature, rainfall, air pressure etc. These developments helped support IMD's two-stage monsoon forecasting system with the first stage forecast given in mid April and an updated second stage forecast given at the end of June.

Kumar et al. (Kumar et al. 2021) conducted a comparative analysis of LSTM models trained using ground-based IMD rainfall data and satellite data from the Tropical Rainfall Measuring Mission for Indian summer monsoon rainfall. They showed correlation coefficients between the observed and predicted rainfall of 0.67 for 1 day and 0.42 for 2 day lead time, respectively, indicating a reasonable skill in short-range precipitation forecasting. However, they observe that the model's efficiency quickly goes down after 2 days lead time.

Another line of work studies oscillations of weather elements (such as dry spells and active spells of rain, and periods of high and low pressures and temperatures) in monsoons. Goswami and Xavier (Goswami and Xavier 2003) showed that it is possible to predict dry spells (i.e., periods of scanty rains) in monsoons with a significantly higher lead time of 18 days as compared to 10 days for active spells (i.e., periods of abundant rains). Goswami et al. (Goswami et al. 2006) found a physical mechanism linking Atlantic climate and monsoon, discovering the influence of Atlantic weather oscillations on variability of Indian monsoon on both interdecadal and interannual time scales.

Rao et al. (Rao et al. 2022) evaluated the skill of NCEP-NWP rainfall forecasts for a few flood-prone basins in India. They observed that as the forecast horizon increased from 1 to 5 days, the bias in rainfall estimation shifted from overestimation to underestimation. They proposed bias correction using a simple multiplicative factor that helped them reduce the mean squared error.

Globally, there has also been an advent of models using sophisticated architectures such as Autoformer (Wu et al. 2021), GraphCast (Lam et al. 2023) and ClimaX (Nguyen et al. 2023). Graphcast uses high frequency reanalysis data from across the globe to predict weather upto 10 days in advance, and outperforms HRES forecasts on all error metrics. Climax is a transformer-based, foundation weather and climate model that is pre-trained on climate simulations from CMIP6 (Eyring et al. 2016). Once pre-trained, the model can be used for multiple climate and weather tasks, where it outperforms all existing data-driven baselines including the Integrated Forecasting System (IFS) from ECMWF (Wedi et al. 2015). While both these models use reanalysis data as the ground truth, we rely on IMD data. It is well known that reanalysis datasets are not fully representative of the surface weather observations in India (Kishore et al. 2016).

Our work adds to the growing literature on using ML for short-term rainfall prediction, primarily using historical IMD data, and benchmarking against NWP forecasts.

2. Data and Experiments

a. Data Sources

1. **IMD Ground Truth**: We use daily gridded precipitation data obtained from IMD spanning the period from 1901 to 2023, at a spatial resolution of $0.25° \times 0.25°$ (Pai et al. 2014). At this resolution, the geographical extent of India is discretized into 12,422 grids. We also use gridded data at $1° \times 1°$ resolution (Rajeevan et al. 2008) during the same period, which provides data at 357 grids. This forms the ground truth dataset against which our predictions and other models are compared.

2. **Additional weather variables**: Apart from precipitation, we also use daily atmospheric and land data at 0.25° resolution provided by ECMWF as part of their reanalysis products (Copernicus Climate Change Service 2019). These variables include: horizontal and vertical components of wind at 10m, temperature, soil moisture, cloud cover, vorticity at 850hPa, humidity, and divergence at 700hPa. These are the lower tropospheric pressure levels, indicative of cloud development and rainfall processes. The data is available from 1950 onwards.

3. **NWP forecasts**: We conduct all comparisons against two popular NWP bechmarks, HRES and NCEP forecasts. The HRES daily forecasts are obtained from ECMWF (ECMWF 2021) for all years 2011 onwards, at a resolution of 0.25°, for both 1 and 3 days into the future. The NWP benchmark from NCEP is obtained from (Copernicus Climate Change Service 2018). It is available more or less on alternate days through the years $2011-2019$, capturing about 60% of the days. Thereafter the data is available daily. The NCEP dataset provides gridded forecasts at 6 hour intervals, for three time periods: 1, 2 and 3 days into the future. The spatial resolution of these forecasts is $1° \times 1°$, with predictions recorded as cumulative values over the subsequent 24 hours at 6-hour intervals. Both the datasets are downloaded only for the JJAS months.

b. Dataset Preparation

We compare the 6 am daily NWP forecasts for lead time of 1 and 3 days with the corresponding deep-learning based forecasts. All datasets are available at $0.25° \times 0.25°$ resolution without any grid mismatch. However, for $1° \times 1°$ comparisons simple linear interpolation is needed to bring all datasets to the same resolution. More details on this alignment process for 1° data are given in the Appendix. For DL-HD and DL-HD+Covariates training, the IMD dataset is partitioned into training and test subsets, covering the periods 1901 to 2011 and 2012 to 2023, respectively. Within the training dataset, individual samples are constructed using a time window approach. Each sample represents d contiguous days of rainfall data from all grid points, serving as the input for the model. The corresponding output is the rainfall data for the $(d+1)^{th}$ day at the same grid points.

c. Experiments

Below we outline how forecasts are generated using different models for lead times of 1 and 3 days.

1. **DL-HD:** We generate forecasts for all n grids across India using historical rainfall data from IMD, utilizing varying lengths of past information, spanning from 3 to 20 days (d). The input dimensions for the models are structured as $n \times d$, capturing the historical rainfall data for all grids over the specified timeframe. The output dimension is n, representing the forecasted rainfall for the subsequent day at each grid point. This is implemented using two separate architectures, LSTM and Autoformer based (Autoformer has a transformer architecture specifically designed for time series data, as described in (Wu et al. 2021)). We train the models from 1901 to 2011, while the test forecasts are generated for the subsequent years from 2012 to 2023. All results are reported only for the Autoformer-based model, as it always outperforms LSTM. This comparison is shown in figure 9.

2. **DL-HD + Covariates:** This is an extension of the above model where we use past d days of precipitation data ($3 \le d \le 20$) from IMD, and past 3 days of reanalysis data. As stated earlier, the additional covariates include wind speed, temperature, soil moisture, cloud cover, vorticity, humidity, and divergence. The choice of these variables is justified later in Section e. The number of past days for reanalysis data were fixed to 3 as we did not observe any significant reduction in errors when incorporating information beyond 3 days. The input dimensions are now structured as $n \times d \times v$, where v is the number of covariates. The output is again n-dimensional, forecasting the rainfall value at each grid.

3. **NWP:** HRES-NWP daily forecasts are available at 0.25° resolution for lead times of 1 and 3 days. These are compared with forecasts made using IMD data at 0.25°. NCEP-NWP forecasts are available at 1° resolution on nearly all alternate days within the test period, and we make comparisons on the days when these forecasts are available with forecasts made using IMD data at 1° resolution. For each grid in the IMD dataset, we identify the best match within the NCEP-NWP re-aligned set and its four adjacent grids based on the criterion of the lowest forecast error. The forecast error is computed for each candidate grid, and the one with the minimum error is taken as the best matching grid. The forecast associated with this identified grid is considered the NCEP-NWP forecast for that specific location.

4. **NWP+:** We combine the HRES-NWP forecasts at the target grid and the 4 neighbouring grids using a deep neural network, which is trained to minimize the error between the forecast and IMD ground truth for the particular grid. The resulting forecast is called HRES-NWP+ prediction. A similar routine is followed for NCEP-NWP forecasts to generate NCEP-NWP+. The models here are trained from 2011 to 2020 and test forecasts are generated for 2021 - 2023. A higher number of surrounding grids does not result in improved forecasts, so we only consider the 4 nearest neighbours.

5. **Ensemble:** We combine the DL-HD + Covariates forecasts, and the NWP forecasts of the 5 grids, to generate an ensemble forecast for each grid. This is done using a deep neural network. The models here are trained from 2011-2020, and forecasts are generated for 2021-2023.

6. **Persistence:** This is a naive forecast which estimates the rainfall on day $d+1$ and the average of rainfall in days $d+1, d+2$ and $d+3$ as the observed rainfall on day d for each grid. This is reported for the period 2012-2023.

d. Loss function

Peak biased loss function for training: Since training on rainfall data is a regression problem, mean squared error (MSE) is the default choice of the loss function. However, using MSE as the loss function often resulted in generated forecasts that were almost constant, resembling the long-term mean of past rainfall. Moreover, to motivate the fact that error of underestimation is more important than the error of overestimation and to be able to capture heavy rainfall better, we assign different exponents to the two errors to obtain the resulting average 'peak-biased' loss function (L). Let $(r_t, t = 1, \ldots, N)$ denote the observed rainfall and $(\hat{r}_t, t = 1, \ldots, N)$ the prediction over N time periods. Then,

$$L = \frac{\sum_{t=1}^{N} \mathbf{I}(\hat{r}_t < r_t)|r_t - \hat{r}_t|^{1.5} + \mathbf{I}(\hat{r}_t > r_t)|\hat{r}_t - r_t|}{N} \tag{1}$$

where $\mathbf{I}(\cdot)$ denotes the indicator function.

The choice of exponent 1.5 in (1) is empirical. In our experiments, we observed that higher exponents resulted in substantial overestimation of peaks, while lower exponents led to underestimation that did a poor job of tracking high rainfall values. We observed a similar trend in the Mean Squared Error (MSE) on the test data, which aligns with the loss function L used for training. See Tables 2 and 3 for details.

e. Feature selection

The input variables for the model are chosen greedily using one feature at a time. We start by considering one variable at a time as input to predict the rainfall in the future, and select the variable with minimum forecasting error as per our loss function. We proceed iteratively. Given that a few variables have been selected at a given stage, we observe the forecasting error resulting from an additional variable, and select the one that gives maximum improvement among all additional variables. The sequential selection process with the errors recorded is outlined in Table 1. It can be seen that past precipitation turns out to be the variable with highest predictive power, followed by cloud cover, vorticity, humidity, soil mositure, wind, and temperature.

TABLE 1: Peak-biased loss ($mm^{1.5} + mm$) recorded for different combinations of input variables using a greedy approach. In each row, we consider the combination of variables indicated in the first column and report the loss when another variable corresponding to the different columns is added to the bucket.

Variables	**Cloud Cover (CC)**	**H Wind (HW)**	**V Wind (VW)**	**Temperature (T)**	**Humidity (H)**	**Soil Moisture (SM)**	**Vorticity (Vo)**
Precipitation (Ppt)	**19.95**	21.13	21.48	21.01	20.71	20.52	20.39
Ppt + CC	-	19.63	19.88	20.20	20.01	**19.38**	19.47
Ppt + CC + Vo	-	19.73	19.71	19.36	**19.07**	19.33	-
Ppt + CC + Vo + H	-	18.91	18.96	18.99	-	**18.56**	-
Ppt + CC + Vo + H + SM	-	**18.42**	18.59	18.53	-	-	-

f. Model Configuration and Training

All experiments are conducted in Python, utilizing the TensorFlow and Keras libraries. DL-HD and DL-HD+Covariates forecasts are made using LSTM and transformer architectures. Both of these are recurrent models and we use past d days of v variables as input, where d ranges from 3 to 20. Both models are designed to have nearly the same number of parameters depending roughly linearly on d, with the number of parameters being approximately 200M for $d = 12$. The transformer architecture is adopted from (Wu et al. 2021). Since there are fewer data points, the models for NWP+ and Ensemble are trained using smaller feed-forward neural networks, with 2 hidden layers. All models are trained using the Adam optimizer, optimizing the peak-biased loss specified in (1). More details on the architecture of these models are given in Table B1 in the Appendix. To ensure robustness of our models, each experiment is conducted across 10 independent runs, employing randomly generated seeds to initialize neural network parameters differently. Performance metrics are reported on the average prediction obtained from these 10 runs.

3. Results

The predictions based on the models specified in Section c are compared with the ground truth daily rainfall data from IMD. Below we present results corresponding to IMD data at $0.25° \times 0.25°$ resolution. The results for $1° \times 1°$ IMD data are given in the Appendix.

1. *Comparison for entire India:*

   The average peak-biased loss over India for 1-day and 3-day forecasts at 0.25° resolution are presented in Tables 2 and 3, respectively. We also compare the spatial distribution of prediction skill for 1 and 3 day prediction on July 15, 2022 in Figures 1 and 2. We make the following observations:

   (a) DL-HD+Covariates outperforms other models significantly for both 1-day and 3-day forecasts.

   (b) Persistence estimates record the highest errors, being 39.36% worse than DL-HD+Covariates 1-day forecasts and 69.52% worse for 3-day forecasts. HRES-NWP, while performing better than perssitence, still reports higher errors compared to DL-HD+Covariates: 21.98% for 1-day forecasts and 27.21% for 3-day forecasts.

(c) Pooling surrounding grids to form NWP+ slightly reduces errors compared to raw HRES-NWP forecasts. However, these forecasts still report higher errors than DL-HD+Covariates forecasts by about 21.33% for 1-day and 25.07% for 3-day predictions.

(d) The ensemble NWP+DL-HD+Covariates exhibits improved performance over NWP alone, and is comparable with DL-HD+Covariates forecasts in most grids. However, it reports slightly higher overall errors, approximately 3.95% worse than DL-HD+Covariates for 1-day forecasts and 11.07% worse for 3-day forecasts.

(e) To evaluate long-term memory in historical data, we conducted experiments using contexts ranging from 3 to 20 days. The detailed experiment and results are given in Section b.

2. *Comparison for key cities:* We also analyzed the model performance separately for 20 of the most populated cities spread across India. The average peak-biased loss on these cities is shown in Table 4 for 1-day forecasts, and in Table 5 for 3-day forecasts. The last row in the two figures shows the excess error percentage in the different forecasts compared to DL-HD+Covariates forecast. We make similar observations here as for whole of India. Figures 3 to 5 graphically compare the different forecasts with the ground truth for cities Mumbai, Ahmedabad and Chennai, for the months of July and August in 2022, for 1-day forecasts. Similar comparisons for 3-day forecasts are shown in Figures 6 to 8. It is clear from the figures that DL-HD+Covariates forecasts consistently outperform other methods in tracking actual rainfall.

a. Additional performance comparisons

In this section we perform additional comparisons between different predictors related to predicting heavy rainfall and generating false alarms for heavy rainfall. We also measure the correlation of different predictors with observed rainfall. Recall that $(r_t, t = 1, \ldots, N)$ denotes the observed rainfall and $(\hat{r}_t, t = 1, \ldots, N)$ the prediction over N time periods.

TABLE 2: Average peak-biased loss ($mm^{1.5} + mm$) and MSE (mm^2) for 1-day forecasts across all the grids in India. Naive persistence estimates have the largest error. HRES-NWP performs significantly better than persistence, yet it remains outperformed by DL-HD. Pooling NWP grids together to form NWP+ gives modest improvement over vanilla NWP. Adding other weather variables along with precipitation gives the lowest errors. The ensemble NWP + DL-HD + Covariates shows significant improvement over NWP alone, although it falls slightly short of DL-HD+Covariates.

Model	**Peak-biased Loss**	**%age Higher Error than DL-HD**	**MSE (mm^2)**
DL-HD + Covariates	18.24	-	268.59
DL-HD	20.90	14.58	312.11
HRES-NWP	22.25	21.98	356.97
HRES-NWP+	22.13	21.33	344.18
Ensemble	18.96	3.95	294.25
Persistence	25.42	39.36	448.10

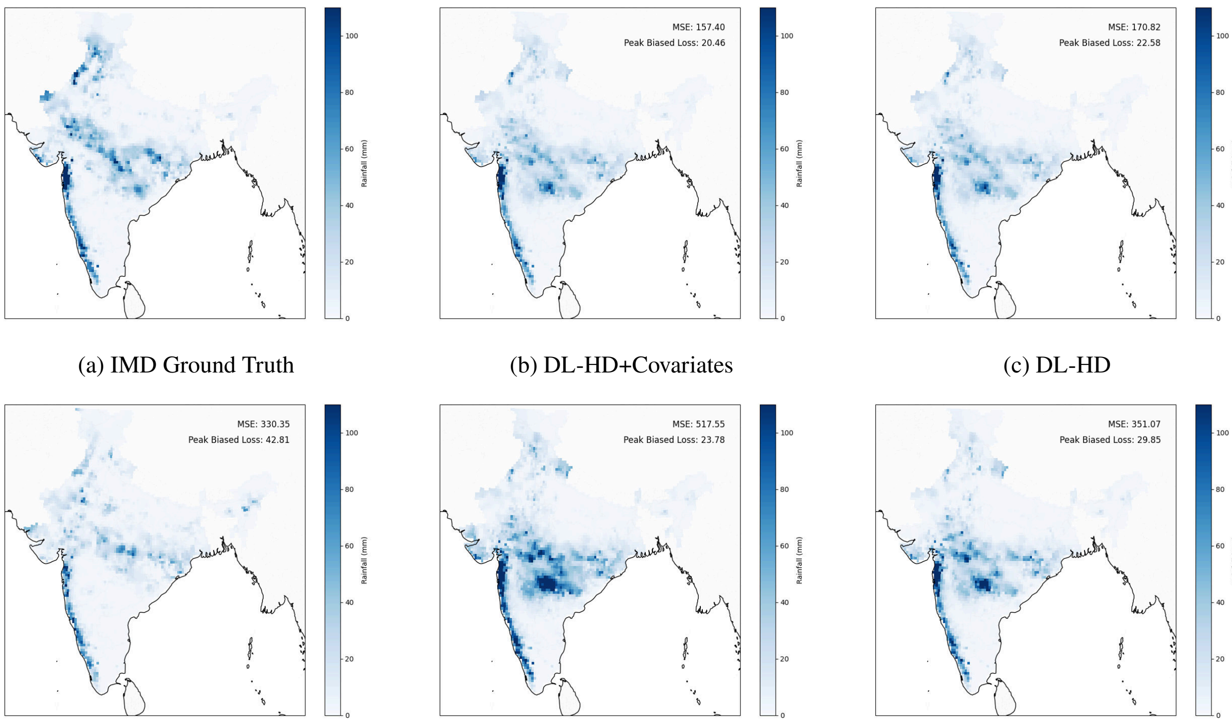
TABLE 3: Average peak-biased loss ($mm^{1.5} + mm$) and MSE (mm^2) for 3-day cumulative forecasts across all the grids in India. The trends are similar to 1-day forecasts. Naive persistence estimates have the largest error, closely followed by NCEP-NWP. HRES-NWP performs significantly better than NCEP-NWP, yet it remains outperformed by DL-HD. Pooling NWP grids together to form NWP+ slightly improves over vanilla NWP. Adding other weather variables along with precipitation gives the lowest errors. The ensemble NWP + DL-HD + Covariates shows significant improvement over NWP alone, although it falls slightly short of DL-HD + Covariates.

Model	**Peak-biased Loss**	**%age Higher Error than DL-HD**	**MSE (mm^2)**
DL-HD + Covariates	67.28	-	2878.52
DL-HD	81.87	21.69	3752.44
HRES-NWP	85.59	27.21	4486.25
HRES-NWP+	84.15	25.07	3884.24
Ensemble	74.73	11.07	3019.81
Persistence	114.06	69.52	8300.21

1. *Heavy rainfall predictor (HRP):* This is calculated as the fraction of times predicted rainfall exceeds a value L when the actual rainfall exceeds a higher value H. A higher HRP value indicates that the model is effective in providing warnings for heavy rainfall. This is important in providing early warnings to people and authorities to prepare for contingencies related to high rainfall. For 1-day forecasts we use $L = 10, H = 30$, and for 3-day forecasts, we use $L = 20, H = 60$. These values were chosen as a sufficient number of events were observed in the key cities for these ranges.

$$HRP = \frac{\#(\hat{r}_t > L, r_t > H)}{\#(r_t > H)} \tag{2}$$

Fig. 1: Spatial distribution of 1-day forecasts for 15 July 2022. DL-HD and NWP+DL-HD predictions detect areas with high rainfall with reasonable effectiveness. NWP forecasts tend to overestimate rainfall across almost all parts of India.



(a) IMD Ground Truth (b) DL-HD+Covariates (c) DL-HD

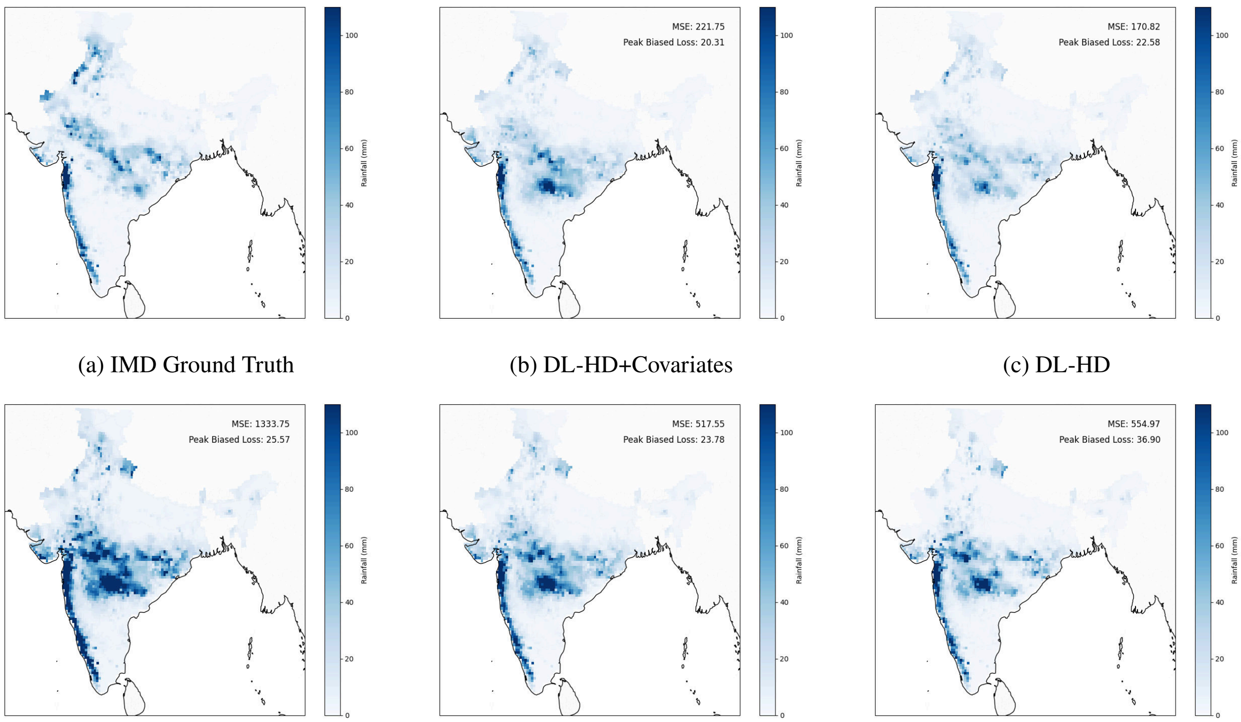
(d) HRES-NWP (e) HRES-NWP+ (f) Ensemble

2. *False alarm rate (FAR):* This is given by the fraction of times predicted rainfall was above a large value H when the actual rainfall was below a small value L. FAR affects the credibility of the model. Higher the FAR, more likely that the warnings of heavy rainfall will be taken less seriously. For 1-day forecasts, we again set $L = 10$ and $H = 30$.

$$FAR = \frac{\#(\hat{r}_t > H, r_t < L)}{\#(\hat{r}_t > H)}. \tag{3}$$

3. *Correlation Coefficient (CC):* As in Kumar et al. (2021), we also study the correlation between the prediction over N time periods and the observed rainfall. A higher absolute value of CC indicates a stronger linear relationship between the predicted and observed time series. Letting $\bar{r} = N^{-1} \sum_{t=1}^{N} r_t$, the CC equals

FIG. 2: Spatial distribution of 3-day forecasts for 15 July 2022. Transformer based predictions detect areas with high rainfall with reasonable effectiveness. NWP forecasts tend to overestimate rainfall across almost all parts of India.



(a) IMD Ground Truth

(b) DL-HD+Covariates

(c) DL-HD

(d) HRES-NWP

(e) HRES-NWP+

(f) Ensemble

$$\frac{\sum_{t=1}^{N}(\hat{r}_t - \bar{\hat{r}})(r_t - \bar{r})}{\sqrt{\sum_{t=1}^{N}(\hat{r}_t - \bar{\hat{r}})^2 (r_t - \bar{r})^2}}. \tag{4}$$

Table 6 displays these metrics for all of India for 1-day as well as 3-day forecasts. Note that all numbers are reported as average of the respective metrics over all the grids.

We make the following observations:

1. At all India level, in case of 1-day forecasts, NCEP-NWP outperforms other models in terms of HRP, detecting about 84% of high rainfall events. This however, is due to its tendency to overestimate rainfall on most days. DL-HD+Covariates and the ensemble model follow closely behind with HRP 82.24% and 82.61% respectively. Persistence reports a poor HRP of 59.11%. The strength of DL-HD and DL-HD+Covariates lies in that they have a high HRP

TABLE 4: Average peak biased loss ($mm^{1.5} + mm$) for 1-day forecasts in grids corresponding to 20 major cities across India. Overall, and in most cities, DL-HD+Covariates outperforms other models by a significant margin. The ensemble model combining NWP and DL-HD+Covariates follows closely behind in the total error, and even performs better than just DL-HD+Covariates in some cities. HRES-NWP performs the poorest, having a 27.77% higher error.

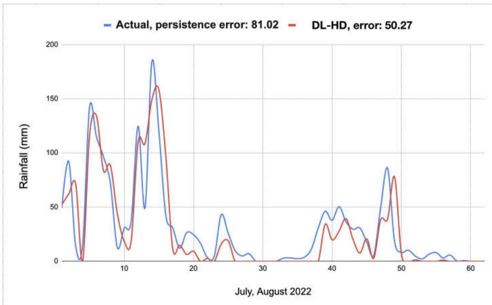
City	**DL-HD+Covariates**	**DL-HD**	**HRES-NWP**	**HRES-NWP+**	**Ensemble**
Ahmedabad	**16.23**	18.46	25.34	23.15	17.32
Bangalore	11.76	12.28	14.67	12.33	**11.12**
Bhopal	**24.35**	26.76	28.92	27.14	23.98
Bhubaneswar	**26.42**	29.11	28.04	27.91	27.74
Chandigarh	18.45	**18.35**	20.42	20.18	19.65
Chennai	**10.33**	11.15	12.40	12.26	12.18
Coimbatore	10.27	10.77	11.55	11.31	**10.09**
Delhi	**7.44**	7.86	12.79	11.77	8.85
Gangtok	**38.71**	41.47	42.29	41.93	39.33
Hyderabad	**18.55**	20.28	26.51	23.14	20.08
Indore	11.22	11.41	15.65	15.41	**10.40**
Kochi	19.26	22.77	26.13	25.42	**19.07**
Kolkata	**35.96**	37.58	41.75	40.59	36.86
Lucknow	11.86	**11.72**	16.46	16.11	11.99
Mumbai	**42.48**	47.50	66.13	63.81	45.05
Patna	12.71	13.33	15.58	15.51	**12.46**
Pune	**15.11**	16.85	25.46	25.29	16.78
Raipur	**26.38**	27.51	29.53	28.17	28.47
Shimla	**9.62**	11.77	12.21	11.61	9.82
Vishakhapatnam	23.49	26.75	27.21	27.18	**22.53**
Total Error	**390.60**	423.69	499.36	481.01	402.77
%age higher	0	8.38	27.77	23.05	3.08

and at 31.11% and 27.16%, respectively, a low FAR. HRES-NWP and NCEP-NWP on the other hand have high FAR of 41.2% and 48.14% respectively. Persistence reports a FAR of 46.24%.

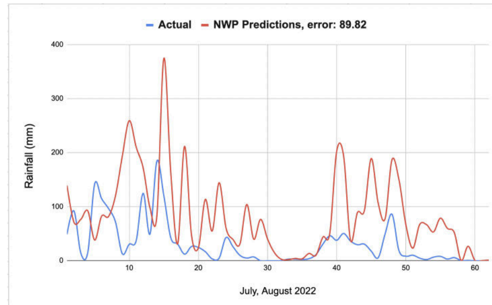
2. For 3-day forecasts, DL-HD+Covariates reports a 74.8% HRP compared to 71.23% and 67.85% HRP by HRES-NWP and NCEP-NWP respectively, and 58.54% HRP by persistence. The ensemble represents a significant improvement over NWP alone, with HRP of 74.78%.

3. DL-HD+Covariates has the highest overall correlation with IMD data, for both 1-day (0.82) as well as 3-day forecasts (0.64). HRES-NWP reports much lower respective correlations of (0.69 and 0.59), persistence reports respective correlations (0.49, 0.57).

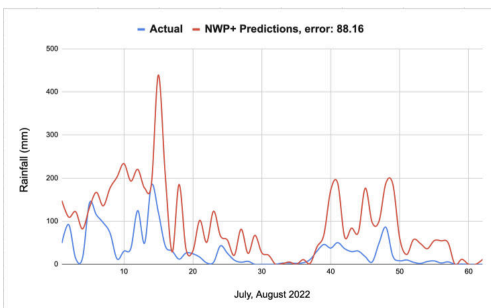
FIG. 3: 1-day forecasts for Mumbai in July and August 2022. DL-HD+Covariates predictions closely track the ground truth, while HRES-NWP predictions tend to over estimate the rainfall. The ensemble is a significant improvement over NWP alone, and can be seen to capture most of the high rainfall events during this period.



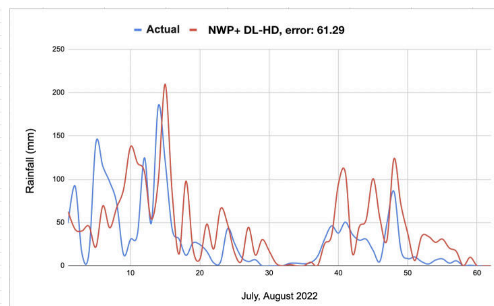
(a) DL-HD + Covariates vs IMD



(b) HRES-NWP vs IMD



(c) HRES-NWP+ vs IMD

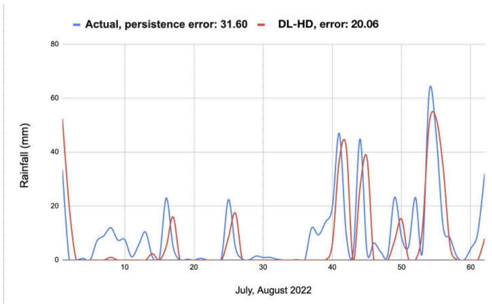


(d) Ensemble vs IMD

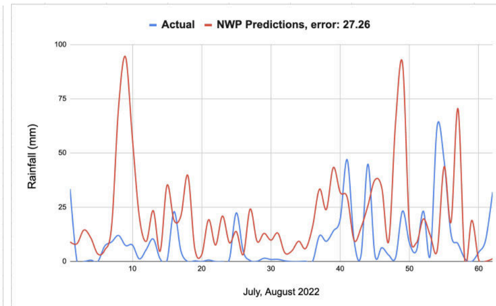
b. Spatio-temporal information in rainfall observations across India

To check for long-term memory in historical data, we conducted experiments for contexts ranging from 3-20 days. We observed a pattern of small but consistently decreasing errors with increasing lag as shown in Figure 9. It can also be observed that there are no significant changes in error for information beyond 12 days at an all India level. We also show the effect of this long-term memory for a few key cities in India. We observe that the improvement in forecasts with historical context is more significant for landlocked cities (such as Bangalore, Hyderabad, Delhi) as compared to coastal regions (such as Mumbai, Chennai, Kolkata). In Figure 10b, we compare the aggregated landlocked regions with the coastal ones. This further strengthens our observation that improvement in forecasts with historical context is more significant for landlocked regions compared to coastal ones. This also suggests that coastal region prediction accuracy may improve a little if rainfall data from neighbouring regions was available.

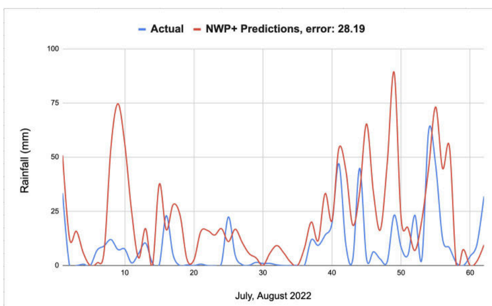
FIG. 4: 1-day forecasts for Ahmedabad in July and August 2022. DL-HD+Covariates closely track the ground truth, capturing most high rainfall events during this period. HRES-NWP predictions consistently overestimate the rainfall. The ensemble performs well in tracking the rainfall but tends to underestimate the precipitation.



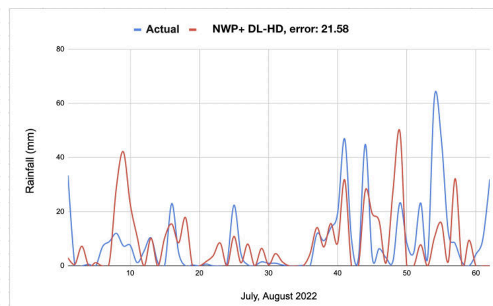
(a) DL-HD + Covariates vs IMD



(b) HRES-NWP vs IMD



(c) HRES-NWP+ vs IMD

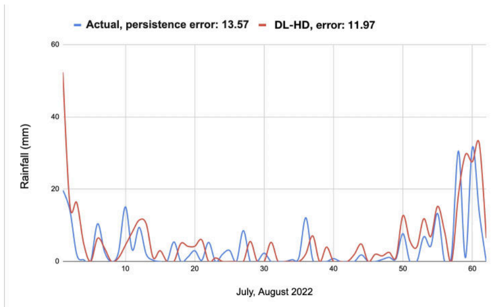


(d) Ensemble vs IMD

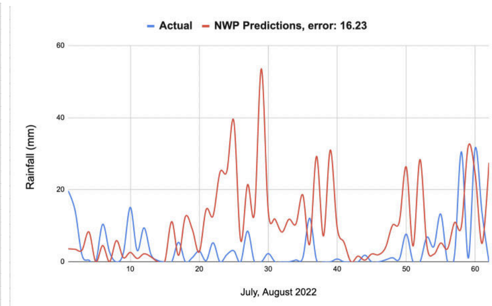
4. Conclusion

In this paper, we compared monsoon rainfall forecasts across India, one day as well as three days in advance at a spatial resolution of $0.25^\circ \times 0.25^\circ$, as well as $1^\circ \times 1^\circ$. We trained a deep neural network using historical rainfall data and other weather variables to arrive at DL-HD+Covariates forecasts. These were compared to persistence based forecasts as well as NWP forecasts obtained from NCEP and ECMWF. We found that forecasts obtained by DL-HD+Covariates are substantially more accurate compared to NWP forecasts as well as predictions based on persistence. We also discussed the improvement in performance of our predictions when they are combined with NWP predictions. We further observed that data up to 20 days in the past is useful in reducing errors of one and three day forecasts, especially when the transformer based learning architecture is used. As noted in the abstract, a key conclusion suggested by our preliminary analysis is that more and diverse data relevant to monsoon prediction, including satellite and radar based data, combined

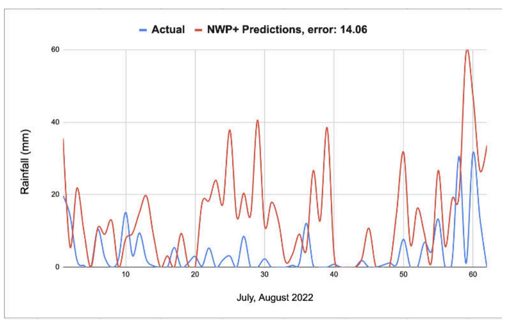
FIG. 5: 1-day forecasts for Chennai in July and August 2022. DL-HD+Covariates closely track the IMD ground truth. HRES-NWP predictions consistently overestimate the rainfall. The ensemble performs well in tracking the rainfall but tends to underestimate the precipitation, resulting in a larger loss than DL-HD+Covariates



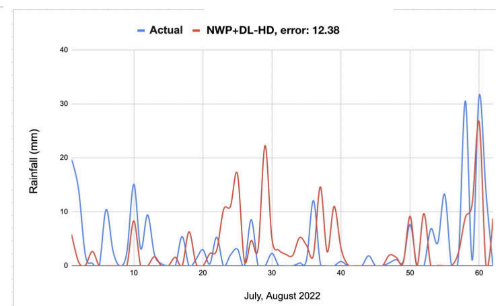
(a) DL-HD + Covariates vs IMD



(b) HRES-NWP vs IMD



(c) HRES-NWP+ vs IMD



(d) Ensemble vs IMD

with carefully selected neural network architecture is likely to substantially improve upon existing NWP forecasts.

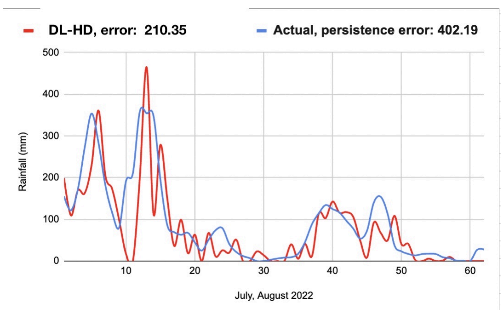
TABLE 5: Average peak biased loss ($mm^{1.5} + mm$) for 3-day forecasts in grids corresponding to 20 major cities across India. Overall, and in most cities, DL-HD+Covariates outperforms other models by a significant margin. The ensemble model combining NWP and DL-HD+Covariates follows closely behind in the total error. HRES-NWP performs the poorest, having a 28.51% higher error.

City	**DL-HD+Covariates**	**DL-HD**	**HRES-NWP**	**NWP+**	**Ensemble**
Ahmedabad	**58.89**	65.22	66.29	65.96	59.88
Bangalore	40.39	44.73	49.11	48.47	**38.80**
Bhopal	**72.25**	79.16	82.43	81.57	75.49
Bhubaneswar	**56.76**	69.46	74.03	74.14	64.32
Chandigarh	**45.87**	50.25	55.27	53.34	48.19
Chennai	**33.53**	43.10	51.54	49.78	40.22
Coimbatore	**42.15**	47.36	50.96	50.48	46.14
Delhi	**32.16**	32.53	40.53	39.81	32.11
Gangtok	**79.56**	100.39	112.74	109.56	88.46
Hyderabad	**42.27**	50.27	54.13	54.22	44.91
Indore	57.42	57.29	62.94	60.41	**56.71**
Kochi	**59.49**	69.56	74.21	74.44	62.24
Kolkata	**94.26**	112.73	118.36	115.40	99.18
Lucknow	**30.71**	34.58	53.91	51.76	32.15
Mumbai	**153.56**	201.28	215.68	211.42	166.77
Patna	**29.93**	34.22	41.39	41.14	30.28
Pune	**40.75**	50.45	54.23	52.61	46.86
Raipur	**64.66**	73.29	81.64	80.11	70.13
Shimla	**22.42**	31.58	34.75	34.49	21.94
Vishakhapatnam	**72.48**	80.56	76.50	75.93	76.54
Total Error	**1130.85**	1331.49	1449.77	1423.93	1223.65
%age higher	0	17.62	28.51	26.23	6.37

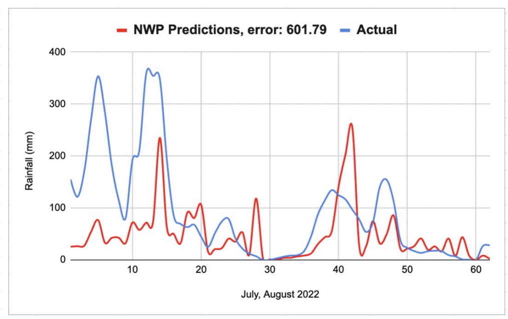
TABLE 6: Comparison of performance metrics for 1-day and 3-day forecasts for entire India. Transformer-based deep learning forecasts do a good job of tracking heavy rainfall while also maintaining a low false alarm rate.

Model	**1-day**			**3-day**		
	FAR (%)	**HRP (%)**	**Correlation Coefficient**	**FAR (%)**	**HRP (%)**	**Correlation Coefficient**
DL-HD + Covariates	**27.16**	82.24	**0.82**	37.41	**74.80**	**0.64**
DL-HD	31.11	80.77	0.75	40.66	74.12	0.62
HRES-NWP	41.20	81.86	0.69	41.39	71.23	0.59
NWP+	39.55	**83.02**	0.62	41.26	72.59	0.61
Ensemble	28.23	82.61	0.81	**37.18**	74.78	0.62
Persistence	46.24	59.11	0.49	48.83	58.54	0.57

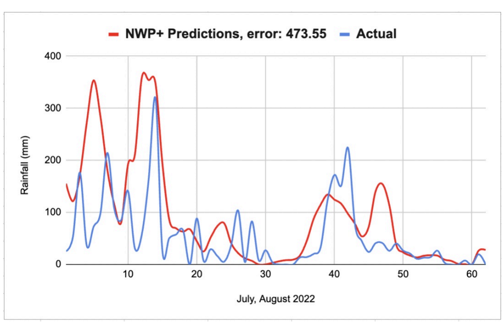
FIG. 6: 3-day forecasts for Mumbai in July and August 2022. None of the forecasts track the IMD ground truth well. However, DL-HD+Covariates capture some of the high and low rainfall events well. HRES-NWP predictions consistently underestimate the rainfall, and we see a significant improvement using HRES-NWP+.



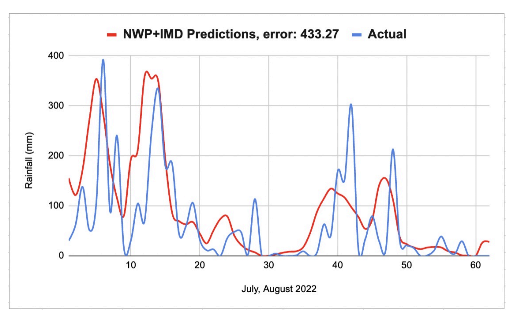
(a) DL-HD+Covariates vs IMD



(b) HRES-NWP vs IMD

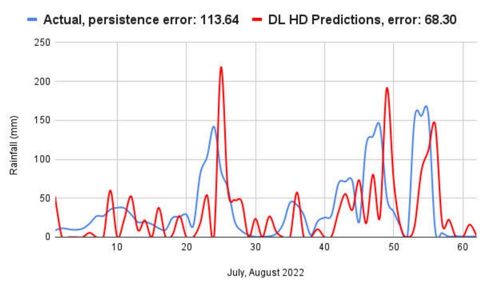


(c) HRES-NWP+ vs IMD

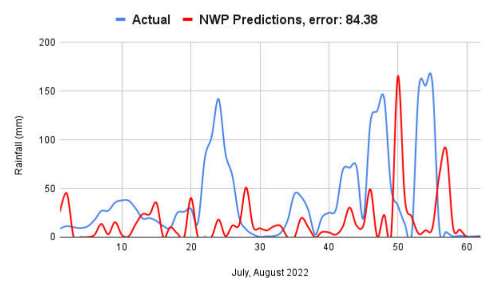


(d) Ensemble vs IMD

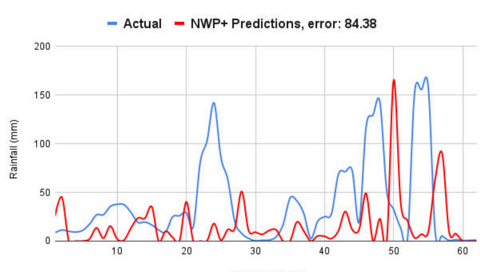
FIG. 7: 3-day forecasts for Ahmedabad in July and August 2022. DL-HD+Covariates track the IMD ground truth closely, and further improve on combining with HRES-NWP forecasts as in the Ensemble, capturing most of the high rainfall events. HRES-NWP predictions too align closely with the IMD ground truth but tend to under estimate the rainfall.



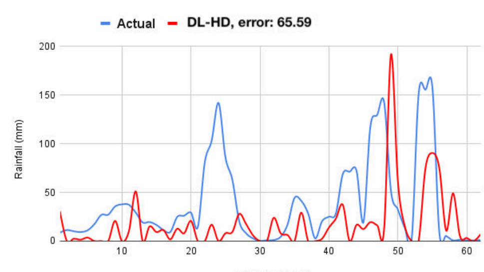
(a) DL-HD+Covariates vs IMD



(b) HRES-NWP vs IMD

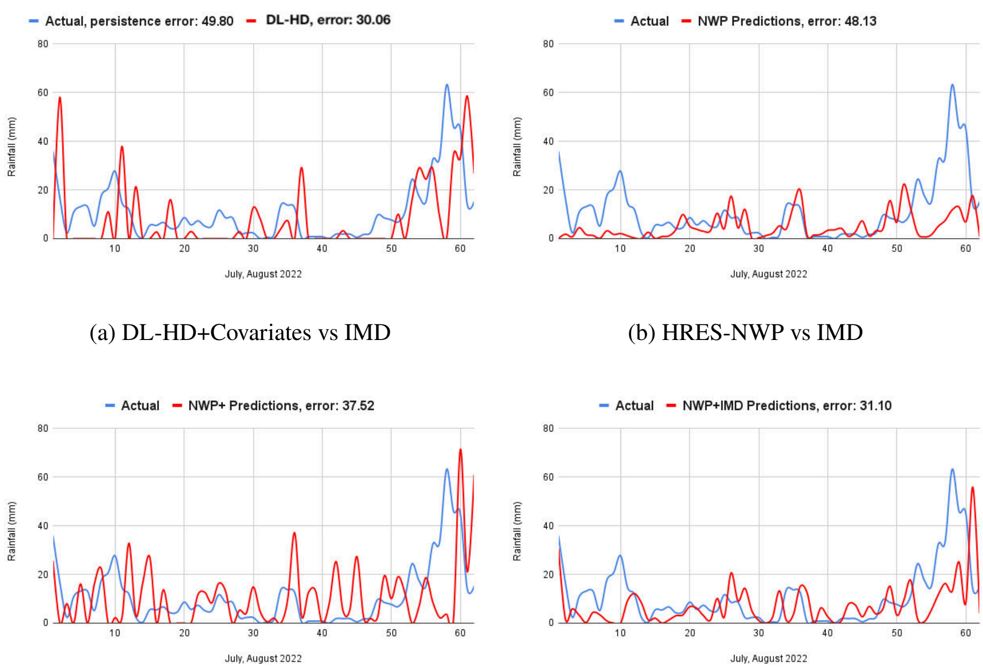


(c) HRES-NWP+ vs IMD



(d) Ensemble vs IMD

Fig. 8: 3-day forecasts for Chennai in July and August 2022. DL-HD+Covariates and Ensemble track the IMD ground truth closely, capturing most of the high rainfall events. Their errors on the peak-biased loss are also comparable. HRES-NWP predictions are not able to capture most rainfall events. We see some improvement with HRES-NWP+, but here too the predictions tend to overestimate the rainfall.



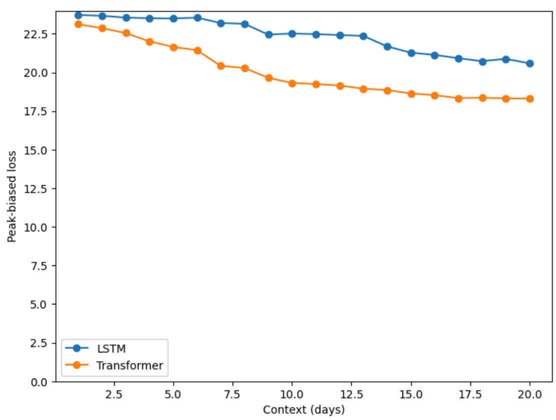
(a) DL-HD+Covariates vs IMD

(b) HRES-NWP vs IMD

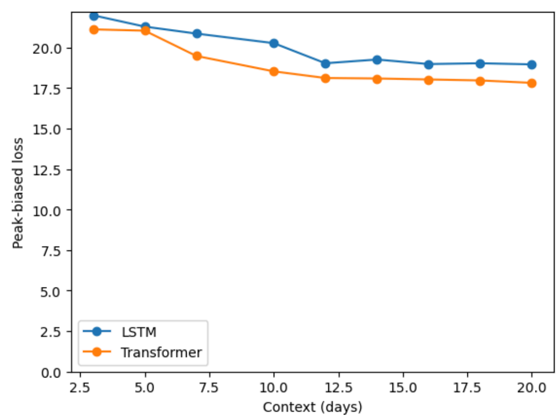
(c) HRES-NWP+ vs IMD

(d) Ensemble vs IMD

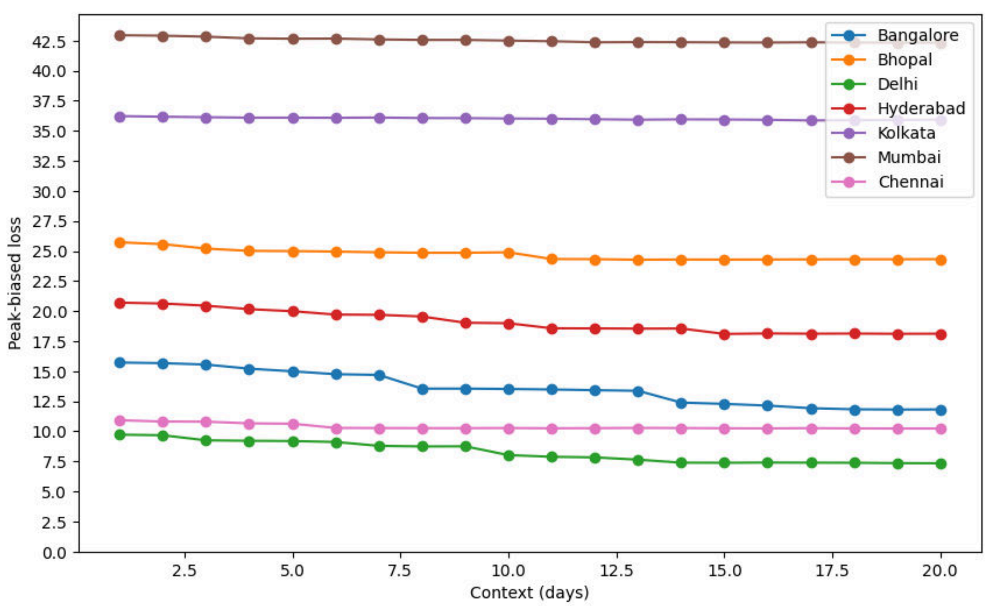
FIG. 9: Average peak-biased loss ($mm^{1.5} + mm$) for 1-day forecasts as a function of number of days of past data (context) used for the forecast. The performance is compared for LSTM and Autoformer architectures. We see a small but consistent decrease in error with context, for both experiments at (a) 0.25° and (b) 1° resolution. In (c), we plot the peak-biased loss as a function of the context in terms of days for a few key cities.



(a)

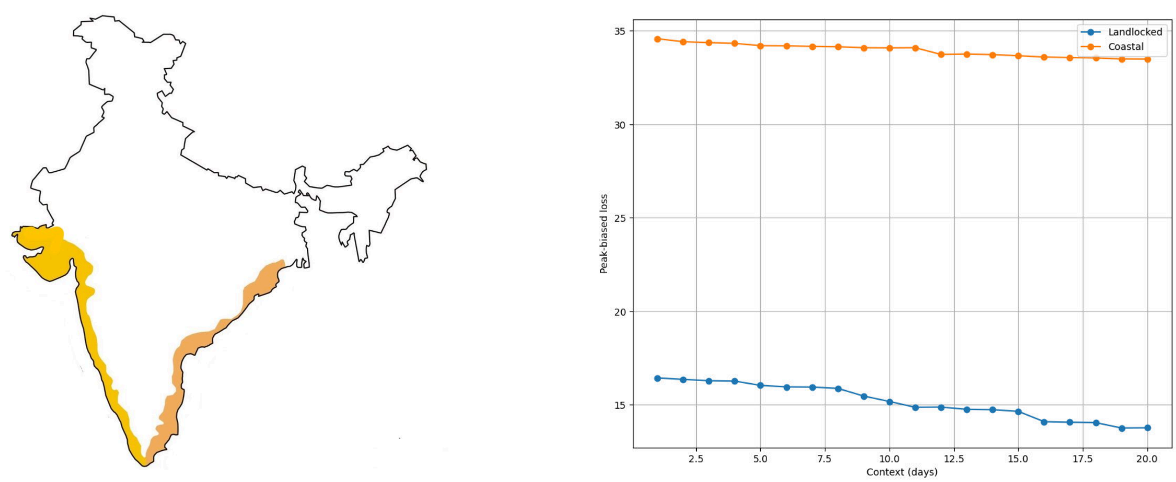


(b)



(c)

Fig. 10: Comparison of average peak biased loss ($mm^{1.5} + mm$) for coastal vs landlocked regions. In (a) the shaded region represents the grids spanning upto 60km from the coastline. (b) compares the error reduction with context for the different regions.



(a) (b)

Acknowledgments. We would like to thank Sandip Trivedi, TIFR for motivating us to work on weather modelling and for related discussions. We thank Debasis Sengupta, ICTS, Rama Govindrajan, ICTS and Partha Mukhopadhyay, IITM for many helpful discussions. We also acknowledge the use of the High-Performance Computing (HPC) facility at TIFR for training large scale models required for this research.

Data availability statement. The data availability statement is where authors should describe how the data underlying the findings within the article can be accessed and reused. Authors should attempt to provide unrestricted access to all data and materials underlying reported findings. If data access is restricted, authors must mention this in the statement. See http://www.ametsoc.org/PubsDataPolicy for more details.

APPENDIX A

Experiments for IMD data at $1^\circ \times 1^\circ$ resolution

a. Dataset Preparation

While the NCEP-NWP dataset shares the same spatial resolution as IMD at $1^\circ \times 1^\circ$, they are offset to each other by 0.5° in both latitude and longitude. IMD reports rainfall for grids whose latitudes and longitudes are multiples of 0.5°, for example the coordinates of a sample grid for IMD could be: [(79.5°, 17.5°), (80.5°, 17.5°), (79.5°, 18.5°), (80.5°, 18.5°)], while NCEP-NWP reports rainfall prediction for grids having integer coordinates. In order to find the best matching grid for IMD from NCEP-NWP, we first align the two by simply shifting all NWP grids by 0.5° to the North-West. To establish a correspondence between each IMD grid and its NCEP-NWP counterpart, we identify the best-matching NCEP-NWP grid among the aligned set and its 4 surrounding grids, selecting the grid with the lowest forecast error.

b. Results

We report the peak-biased loss corresponding to the different models for 1-day forecasts in Table A1. Most trends are similar to those we observed for data at 0.25° resolution, except that NCEP-NWP at $1^\circ \times 1^\circ$ consistently underperforms even persistence based forecasts. The errors for 3-day forecasts are reported in Table A2.

- *Comparison for entire India:* We make the same observations here as for experiments at 0.25° resolution. DL-HD+Covariates outperforms other models significantly for both 1-day and 3-day forecasts. NCEP-NWP forecasts have the highest errors, performing even poorer than persistence estimates, having 43.16% and 65.84% higher errors than DL-HD+Covariates for 1-day and 3-day forecasts respectively. Pooling surrounding grids to form NCEP-NWP+ significantly reduces errors compared to raw NCEP-NWP forecasts. However, these forecasts still report higher errors than DL-HD+Covariates forecasts by about 26.75% for 1-day and 55.85% for 3-day predictions. The ensemble NWP+DL-HD+Covariates exhibits improved performance over NWP alone, and is comparable with DL-HD+Covariates forecasts in most grids. However, it reports slightly higher overall errors, approximately 0.78% worse than DL-HD+Covariates for 1-day forecasts and 10.47% worse for 3-day forecasts.

TABLE A1: Average peak-biased loss for 1-day forecasts across all the grids in India. NCEP-NWP exhibits the highest error, surpassing even the persistence model. The ensemble shows significant improvement over NWP alone, although it falls slightly short of DL-HD+Covariates.

Model	Peak-biased Loss	%age Higher Error
DL-HD+Covariates	16.75	-
DL-HD	18.11	8.12
NCEP-NWP	23.98	43.16
NCEP-NWP+	21.23	26.75
Ensemble	16.88	0.78
Persistence	23.05	37.61

TABLE A2: Average peak-biased loss for 3-day cumulative forecasts across all the grids in India. NWP exhibits the highest error, surpassing even the persistence model. Pooling NWP grids to form NWP+ leads to a substantial improvement in forecasts, yet it remains outperformed by DL-HD. The ensemble NWP+DL-HD shows significant improvement over NWP alone, although it falls slightly short of DL-HD.

Model	Peak-biased Loss	%age Higher Error
DL-HD+Covariates	67.02	-
DL-HD	81.75	21.99
NCEP-NWP	111.14	65.84
NCEP-NWP+	104.46	55.85
Ensemble	74.03	10.47
Persistence	113.36	69.16

- *Comparison for key cities:* We also analyzed the model performance separately for 20 of the most populated cities spread across India. The average peak-biased loss on these cities

TABLE A3: Comparison of performance metrics for 1-day and 3-day forecasts for entire India. Transformer-based deep learning forecasts do a good job of tracking heavy rainfall while also maintaining a low false alarm rate.

Model	1-day			3-day		
	FAR (%)	HRP (%)	Correlation Coefficient	FAR (%)	HRP (%)	Correlation Coefficient
DL-HD + Covariates	**28.86**	79.53	**0.80**	39.02	**73.44**	**0.62**
DL-HD	30.15	78.56	0.78	41.22	73.19	0.60
NCEP-NWP	43.10	83.88	0.66	43.60	69.55	0.57
NCEP-NWP+	40.87	83.02	0.68	42.06	70.94	0.59
Ensemble	31.61	82.83	0.78	**36.12**	72.59	0.60
Persistence	47.15	62.20	0.51	49.76	60.04	0.55

is shown in Table A4 for 1-day forecasts, and in Table A5 for 3-day forecasts. The last row in the two figures shows the excess error percentage in the different forecasts compared to DL-HD+Covariates forecast. We make similar observations here as for whole of India.

APPENDIX B

Neural Network Architecture

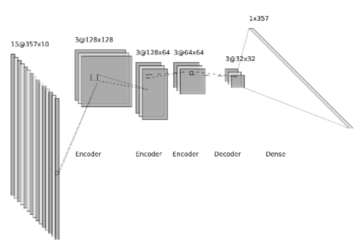


FIG. B1: Neural network architecture used for transformer-based experiments.

References

Copernicus Climate Change Service, 2018: Seasonal forecast daily and subdaily data on single levels. ECMWF, URL https://cds.climate.copernicus.eu/doi/10.24381/cds.181d637e, https://doi.org/10.24381/CDS.181D637E.

TABLE A4: Average peak biased loss for 1-day forecasts in grids corresponding to 20 major cities across India. Overall, and in most cities, DL-HD+Covariates outperforms other models by a significant margin. The ensemble model combining NWP and DL-HD+Covariates follows closely behind in the total error, and even performs better than just DL-HD+Covariates in some cities. NCEP-NWP performs the poorest, having a 46% higher error.

City	DL-HD+Covariates	DL-HD	NCEP-NWP	NCEP-NWP+	Ensemble
Ahmedabad	**18.25**	18.95	32.64	29.88	20.28
Bangalore	**12.22**	13.16	18.64	16.81	13.56
Bhopal	**27.36**	29.88	41.92	41.48	32.11
Bhubaneswar	**29.43**	29.95	45.64	42.10	29.17
Chandigarh	17.78	**17.22**	21.44	19.56	18.28
Chennai	**10.92**	11.13	15.68	13.28	11.64
Coimbatore	**11.47**	12.04	13.23	12.48	11.56
Delhi	10.14	12.24	18.96	14.13	**10.11**
Gangtok	**36.57**	39.67	44.51	41.25	37.75
Hyderabad	**22.50**	22.76	29.24	26.18	23.12
Indore	**14.68**	16.83	20.58	19.54	16.18
Kochi	**20.53**	21.36	31.42	23.44	20.69
Kolkata	**35.95**	36.60	43.55	40.62	36.42
Lucknow	13.40	15.54	23.05	20.37	**12.57**
Mumbai	**45.76**	46.61	89.41	67.84	45.84
Patna	12.23	14.52	18.89	16.22	**11.24**
Pune	15.02	16.93	23.64	22.18	**14.33**
Raipur	26.72	28.11	30.66	30.34	**24.85**
Shimla	9.58	11.28	13.04	12.89	**9.44**
Vishakhapatnam	23.11	26.85	27.88	26.74	**21.65**
Total Error	**413.62**	441.63	604.02	537.33	420.79
%age higher	0	6.77	46.03	29.91	1.73

Copernicus Climate Change Service, 2019: Era5-land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), URL https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac, https://doi.org/10.24381/CDS.E2161BAC.

ECMWF, 2021: Open data. ECMWF, URL https://www.ecmwf.int/en/forecasts/datasets/open-data, https://doi.org/10.21957/OPEN-DATA.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, **9 (5)**, 1937–1958.

TABLE A5: Average peak-biased loss for 3-day forecasts in grids corresponding to 20 major cities across India. Overall, and in most cities, DL-HD+Covariates outperforms other models by a significant margin. The ensemble model combining NWP and DL-HD+Covariates follows closely behind in the total error. NCEP-NWP performs the poorest, having a 47.15% higher error.

City	DL-HD+Covariates	DL-HD	NCEP-NWP	NWP+	Ensemble
Ahmedabad	**57.65**	64.19	78.05	77.12	60.67
Bangalore	**42.19**	46.04	69.02	65.33	50.45
Bhopal	**70.55**	79.82	87.78	86.14	76.31
Bhubaneswar	**54.04**	63.48	78.91	74.52	62.36
Chandigarh	**46.59**	52.18	59.68	58.30	49.63
Chennai	**35.46**	46.27	52.96	50.77	42.93
Coimbatore	**40.56**	47.12	55.64	57.29	42.19
Delhi	**32.05**	34.56	43.27	41.18	36.70
Gangtok	**76.44**	84.58	116.54	107.93	92.16
Hyderabad	**40.95**	43.87	58.76	58.42	49.64
Indore	**54.57**	58.26	68.22	65.73	57.04
Kochi	**62.39**	71.88	86.42	84.83	70.55
Kolkata	**91.62**	101.57	121.34	120.57	97.61
Lucknow	32.11	37.86	65.02	60.88	**32.05**
Mumbai	**151.26**	197.36	259.42	255.65	160.32
Patna	**25.14**	36.78	44.97	45.18	30.63
Pune	**40.25**	53.20	60.81	57.88	45.34
Raipur	**65.00**	74.52	89.05	88.86	69.07
Shimla	21.46	29.84	36.72	36.39	**21.44**
Vishakhapatnam	**65.48**	81.24	94.58	92.43	75.01
Total Error	**1105.76**	1304.62	1627.16	1585.40	1222.50
%age higher	-	17.98	47.15	43.38	10.56

TABLE A6: Comparison of performance metrics for 1-day and 3-day forecasts for entire India. All in all, DL-HD performs significantly better than other forecast methods.

Model	1-day			3-day		
	FAR (%)	HRP(%)	Correlation Coefficient	FAR (%)	HRP(%)	Correlation Coefficient
DL-HD	**33.86**	77.19	**0.65**	35.93	**94.92**	**0.53**
NCEP-NWP	65.59	**81.13**	0.24	37.69	51.86	0.32
NWP+	50.93	72.88	0.42	36.04	82.17	0.38
Ensemble	41.62	79.03	0.33	**31.14**	91.83	0.43
Persistence	44.22	57.28	0.28	34.22	59.62	0.22

Goswami, B. N., M. Madhusoodanan, C. Neema, and D. Sengupta, 2006: A physical mechanism for north atlantic sst influence on the indian summer monsoon. *Geophysical Research Letters*, **33 (2)**.

Table B1: Neural network hyperparameters. For transformer-based experiments, we use default parameters from the Autoformer architecture (Wu et al. 2021). All other hyperparameters were set to their default values.

Parameters	LSTM	NWP	NWP+
Batch size	64	24	24
Hidden layers	400-100-50-357	32-16-1	32-16-1
Activation function	Default-Default-Sigmoid	ReLU-Sigmoid	ReLU-Sigmoid
Epochs	300	100	100
Early stopping	20	10	10

Goswami, B. N., and P. K. Xavier, 2003: Potential predictability and extended range prediction of indian summer monsoon breaks. *Geophysical Research Letters*, **30 (18)**.

IMD, 19-04-2017: Numerical Weather Prediction Guidance. https://mausam.imd.gov.in/responsive/pdf_viewer_css/met1/Chapter-8%20page%20174-190/Chapter-8%20page%20174-190.pdf.

Kishore, P., S. Jyothi, G. Basha, S. Rao, M. Rajeevan, I. Velicogna, and T. C. Sutterley, 2016: Precipitation climatology over india: validation with observations and reanalysis datasets and spatial trends. *Climate dynamics*, **46**, 541–556.

Kumar, B., and Coauthors, 2021: Deep learning based forecasting of indian summer monsoon rainfall. *algorithms*, **9 (11)**, 13.

Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382 (6677)**, 1416–1421.

Nguyen, T., J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, 2023: Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*.

Pai, D., M. Rajeevan, O. Sreejith, B. Mukhopadhyay, and N. Satbha, 2014: Development of a new high spatial resolution (0.25× 0.25) long period (1901-2010) daily gridded rainfall data set over india and its comparison with existing data sets over the region. *Mausam*, **65 (1)**, 1–18.

Prasad, K., and S. Singh, 1988: Large-scale features of the indian summer monsoon rainfall and their association with some oceanic and atmospheric variables. *Advances in atmospheric sciences*, **5 (4)**, 499–513.

Rajeevan, M., J. Bhate, and A. K. Jaswal, 2008: Analysis of variability and trends of extreme rainfall events over india using 104 years of gridded daily rainfall data. *Geophysical research letters*, **35 (18)**.

Rajeevan, M., P. Guhathakurta, and V. Thapliyal, 2000: New models for long range forecasts of summer monsoon rainfall over north west and peninsular india. *Meteorology and Atmospheric Physics*, **73**, 211–225.

Rajeevan, M., D. Pai, R. Anil Kumar, and B. Lal, 2007: New statistical models for long-range forecasting of southwest monsoon rainfall over india. *Climate Dynamics*, **28**, 813–828.

Rajeevan, M. N., 23-12-2023: T.N. rains— Predicting extreme rainfall events using probabilistic forecasts. https://www.thehindu.com/sci-tech/science/predicting-extreme-rainfall-events-using-probabilistic-forecasts/article67666205.ece#.

Rao, G. V., K. V. Reddy, V. Sridhar, R. Srinivasan, N. Umamahesh, and D. Pratap, 2022: Evaluation of ncep-gfs-based rainfall forecasts over the nagavali and vamsadhara basins in india. *Atmospheric Research*, **278**, 106 326.

Wedi, N., and Coauthors, 2015: *The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges*. European Centre for Medium-Range Weather Forecasts.

Wu, H., J. Xu, J. Wang, and M. Long, 2021: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, **34**, 22 419–22 430.