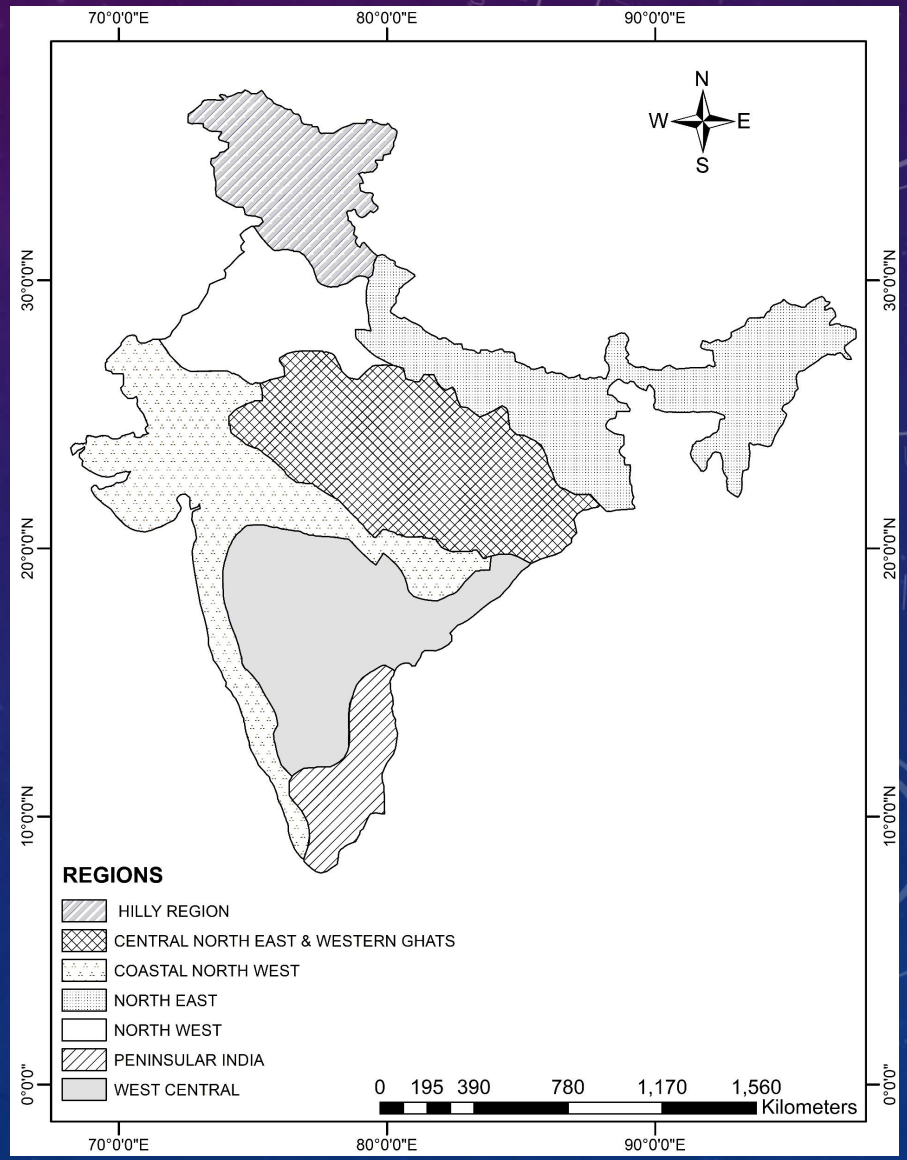
WEATHER FORECASTING USING DEEP LEARNING ON ISMR DATASET

OELP

Ishwar Govind
111901024

ISMR DATASET

- Indian Summer Monsoon Rainfall Dataset
- Contains precipitation of each month from year 1901 – 2015
- 4964 Grid points (whole India)
- Divided into 7 regions
- Values in terms of mm rainfall
- Largest Rainfall Value 6881.4
- Large Scale Climatic Indices Used
  - ENSO - El Niño–Southern Oscillation
  - PDO - Pacific decadal oscillation
  - AMO - Atlantic multidecadal oscillation



MOTIVATION

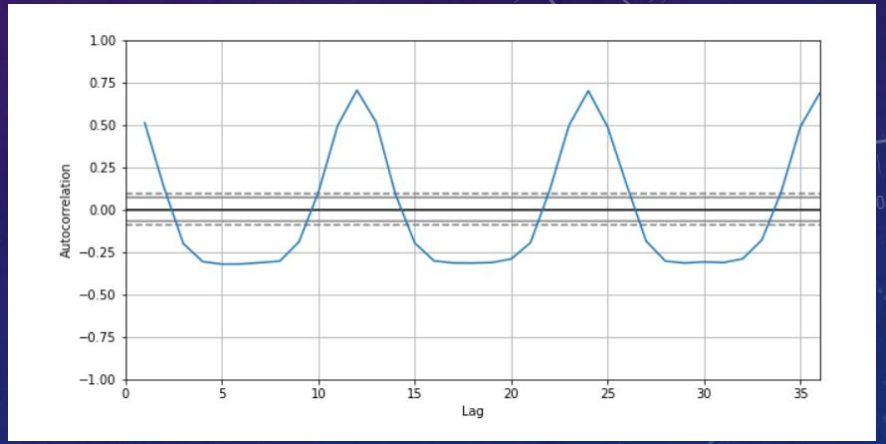
- Improve Current Forecasting framework for Rainfall Prediction in India.
- Main Focus is on Drought and Flood Prediction.
- Current State of the Art in ISMR is done using basic ML and ANN models.
- ConvLSTM used previously for Precipitation Nowcasting in Hong Kong region.
- Deep Learning has shown to work well in forecasting El-Nino/Southern Oscillations.

Shi, Xingjian, et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." Advances in neural information processing systems 28 (2015).

Yoo-Geun Ham, Jeong-Hwan Kim, Eun-Sol Kim, Kyoung-Woon On, Unified deep learning model for El Niño/Southern Oscillation forecasts by incorporating seasonality in climate data,Science Bulletin,Volume 66, Issue 13,2021

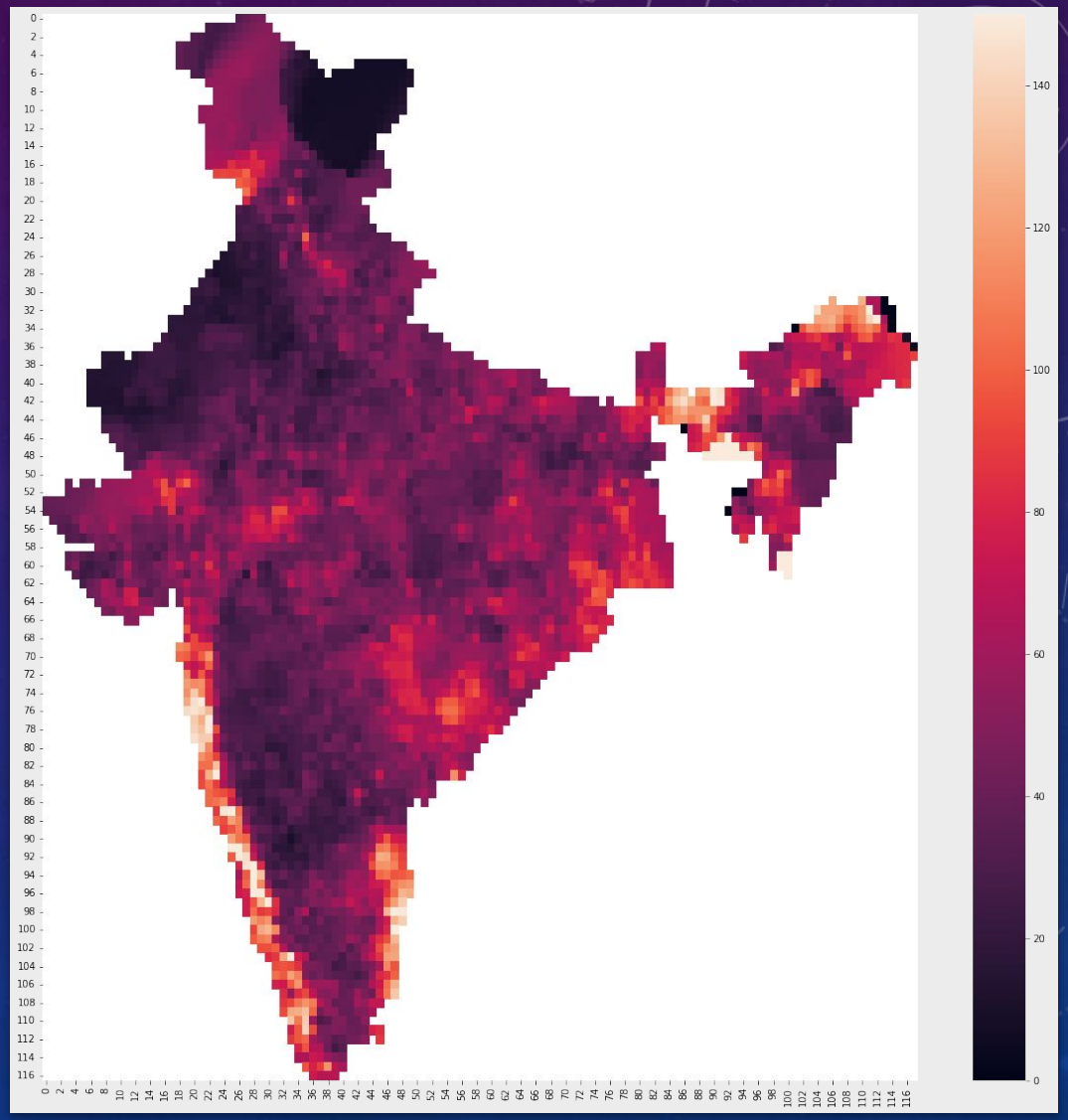
AUTOCORRELATION PLOT

- To find correlation between monthly precipitation values.
- Multiples of 12 showing high correlation.
- Use 12 as horizon as requires less data to train.
- Grid Point used – 2033 (Latitude 22.75, Longitude – 76.25)



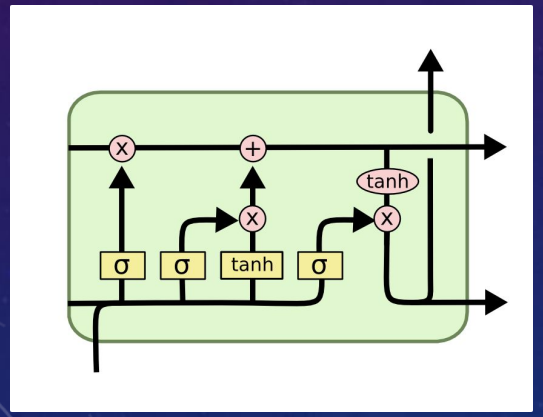
AUTOREGRESSION MODEL

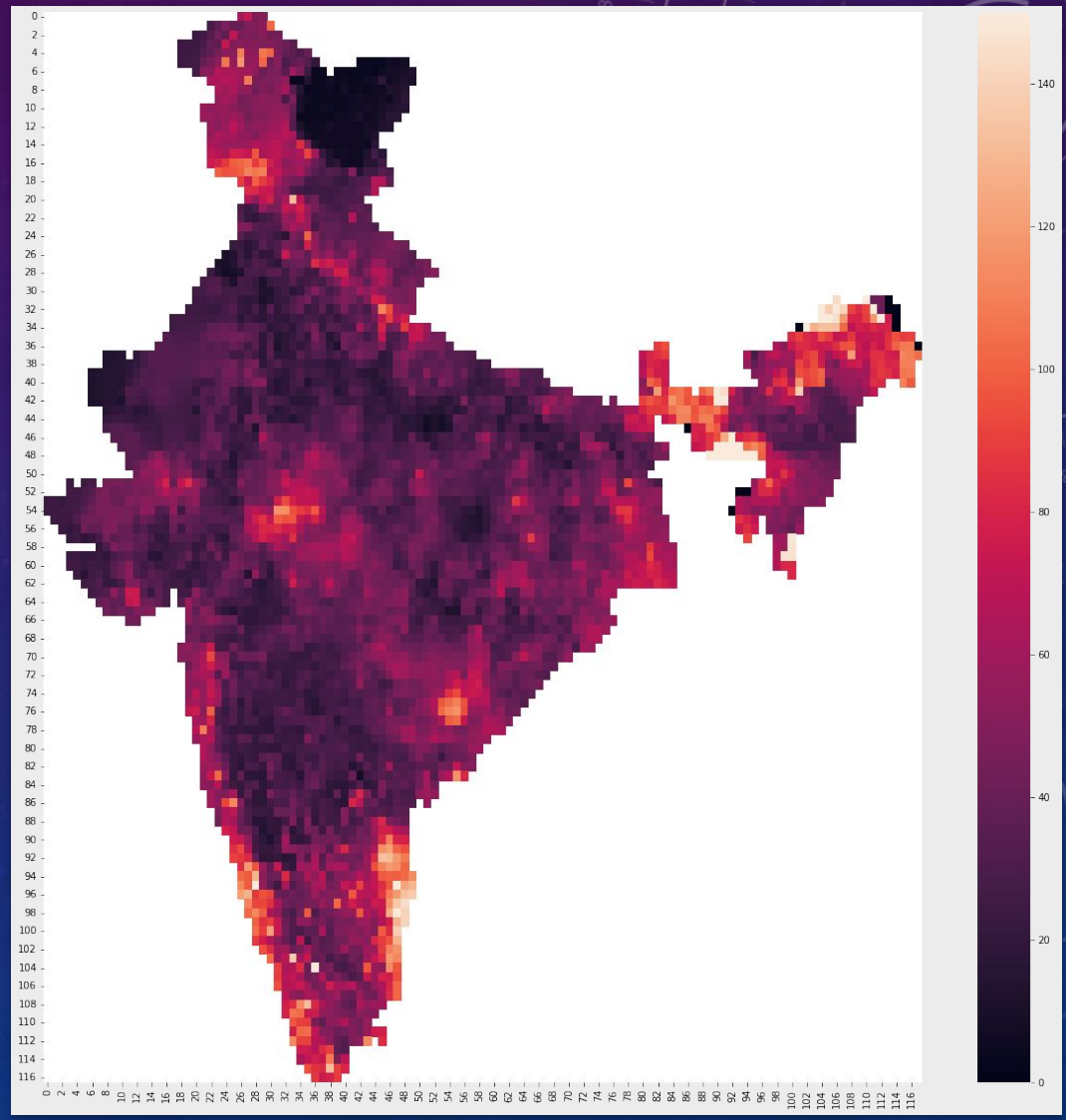
- Time series model that uses observations from previous time steps as input to a regression equation to predict the next step.
- $x_{n+1} = w_0 + w_1 * x_1 + w_2 * x_2 + \ldots.. + w_n * x_n$
- Trained on first 1366 months and tested on the last 14 months
- Error Analysis method (Mean Absolute Error)
- Coastal Regions and North East having highest errors
- Central west and Hilly East least error
- Lesser error means easily learnable pattern in precipitation



LSTM(LONG SHORT-TERM MEMORY)-BASED MODEL

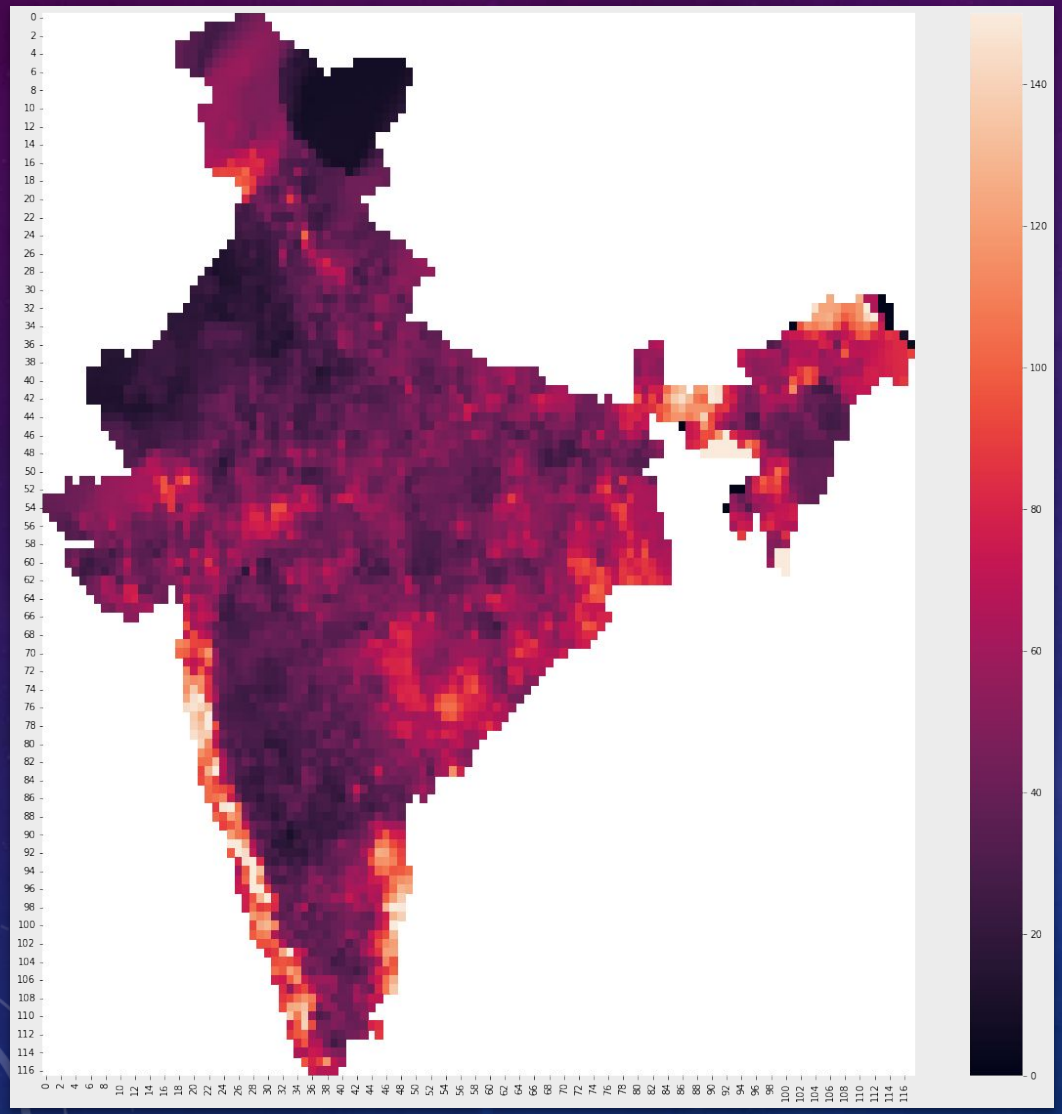
- Max Value of precipitation is 6881.4
- Denormalized Mean Absolute Error.
- Inputs contains 12 month's of vector of length 4964(India grid points) + 3(ENSO,PDO,AMO).
- Better predictions in the eastern part of hilly regions.
- Smaller error when compared to Autoregression.
- Better predictions in central India.

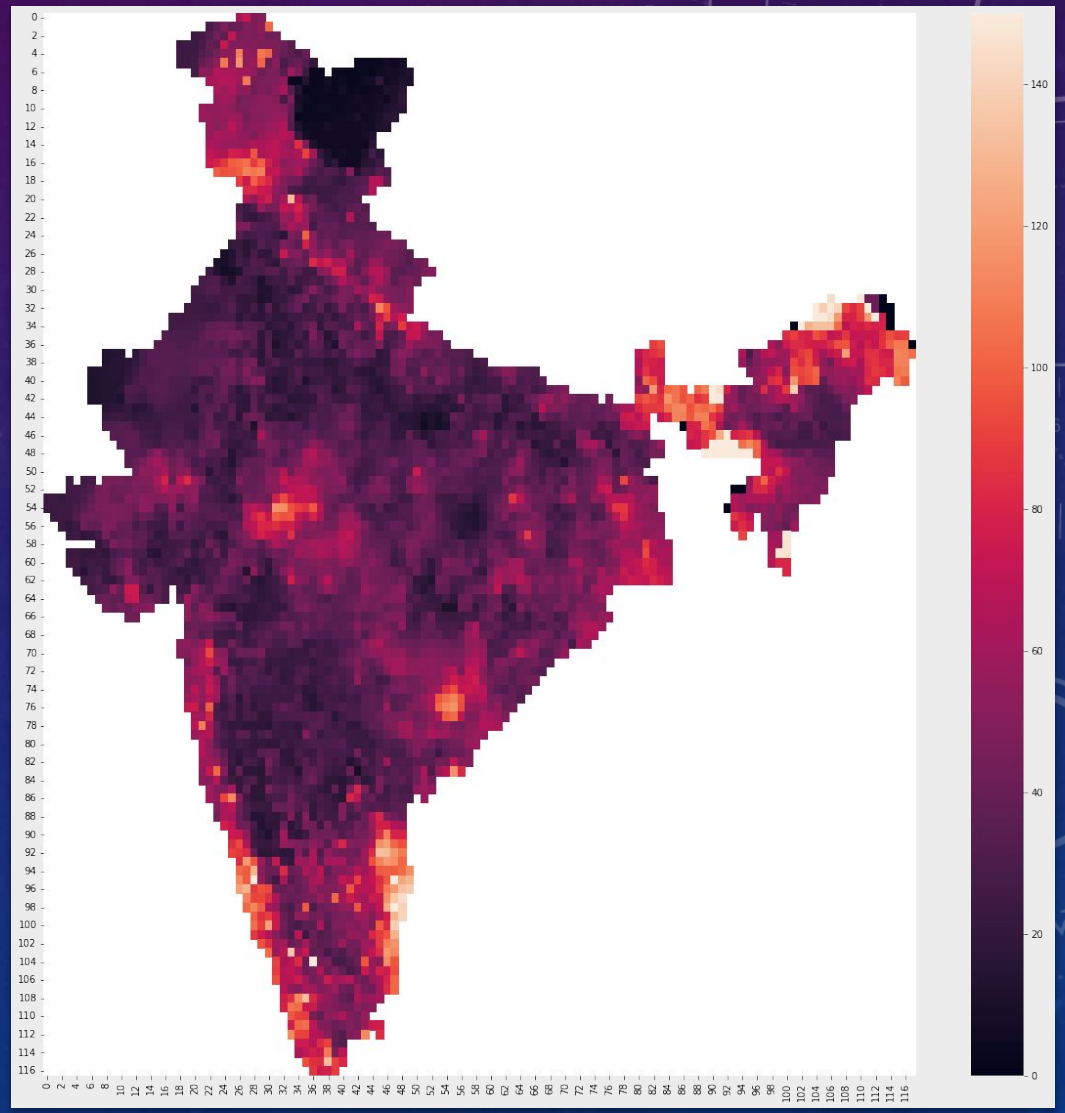




Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

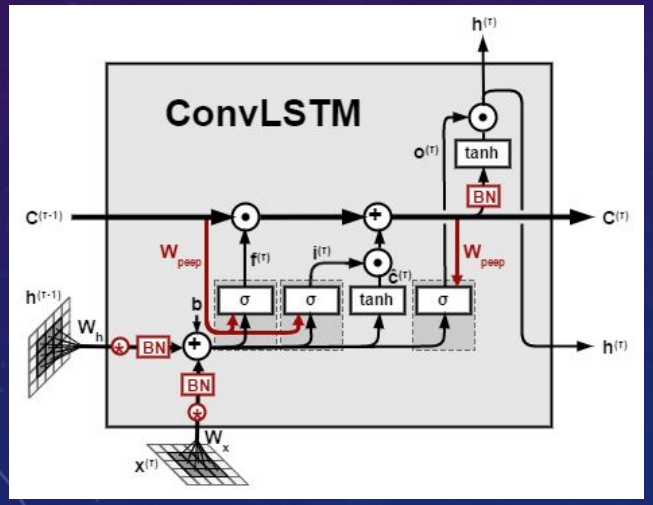
COMPARISON AUTOREGRESSION VS LSTM

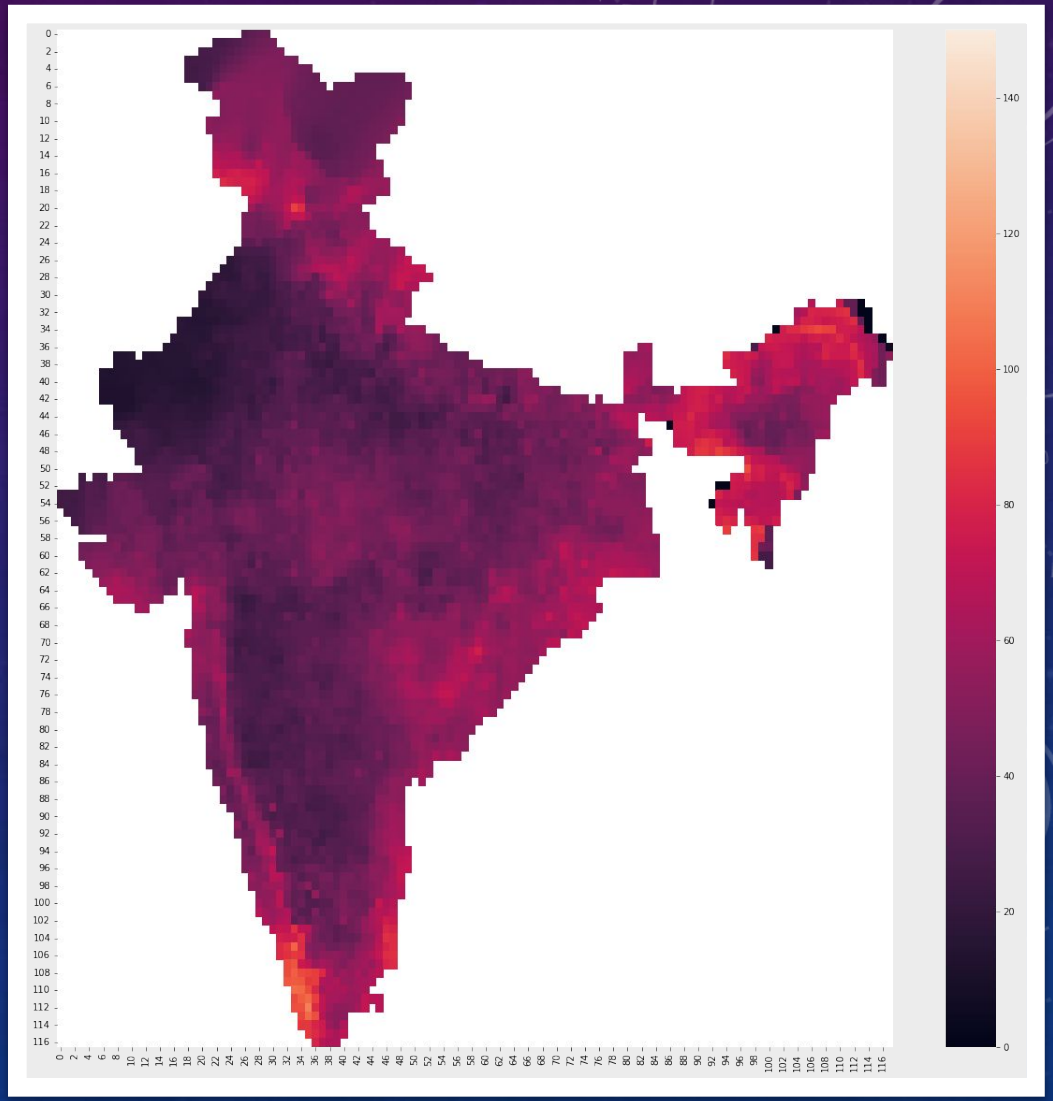




CONVLSTM(CONVOLUTIONAL LSTM)-BASED MODEL

- Lesser error when compared to LSTM and Autoregression.
- Better prediction at central regions and north western region.
- Very high error values at north eastern region and south west coastal regions.
- Higher error values at Hilly regions.





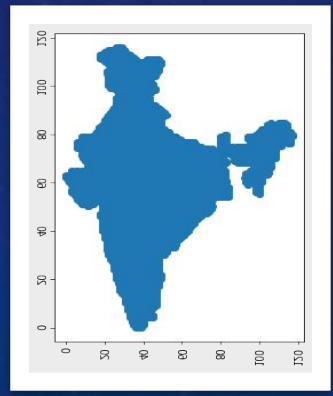
Shi, Xingjian, et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." *Advances in neural information processing systems* 28 (2015).

MEAN ABSOLUTE ERROR OF CONVLSTM WITH CLIPPING FOR DIFFERENT COMBINATIONS OF CLIMATIC OSCILLATIONS

ENSO	PDO	AMO	Error (Grid Loss)
0	0	0	32.257698
0	0	1	17.542728
0	1	0	17.838507
0	1	1	17.993984
1	0	0	16.166445
1	0	1	18.905806
1	1	0	17.787107
1	1	1	18.343975

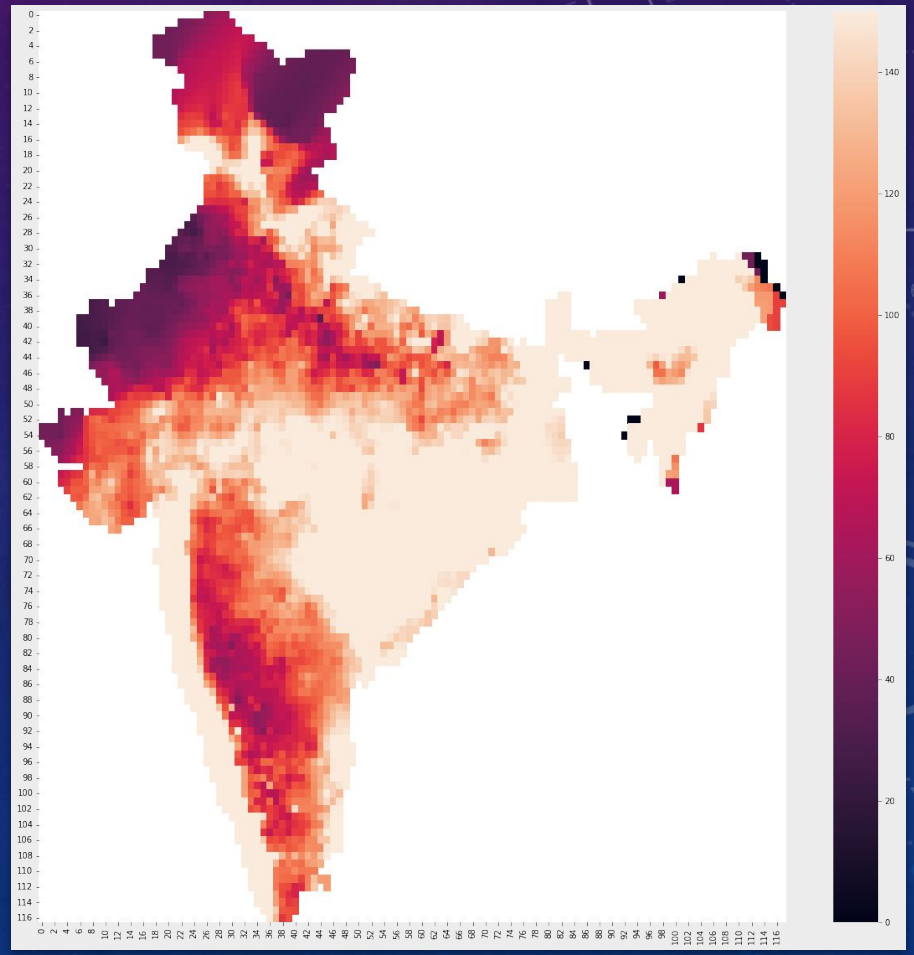
- Output produces a 2D grid of dimensions (117,118)
- Only required to calculate MAE(mean absolute error) of outputs predicted in the region of India.

- Grid Loss - The MAE of the complete 2D array is calculated and multiplied by a mask which only considers the region of India.

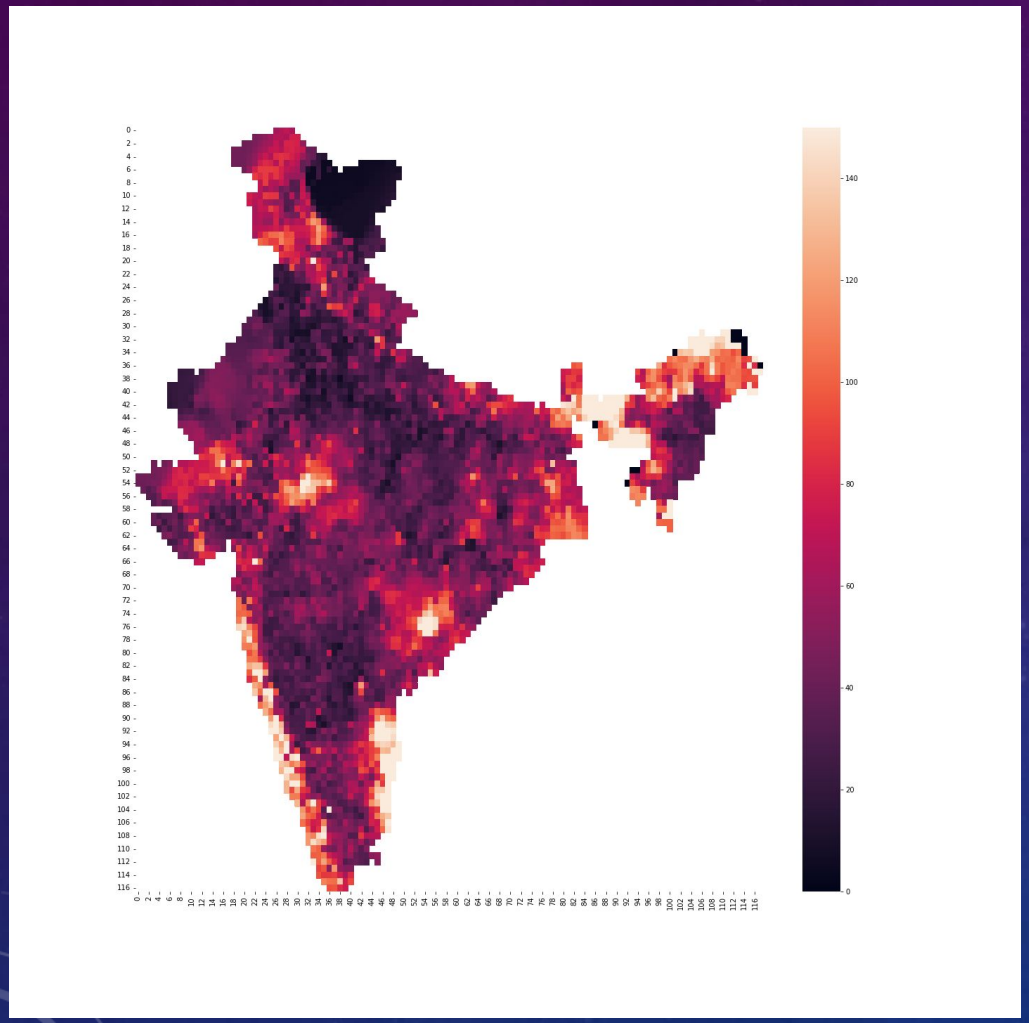


ISSUES OF CONV-LSTM-BASED MODEL

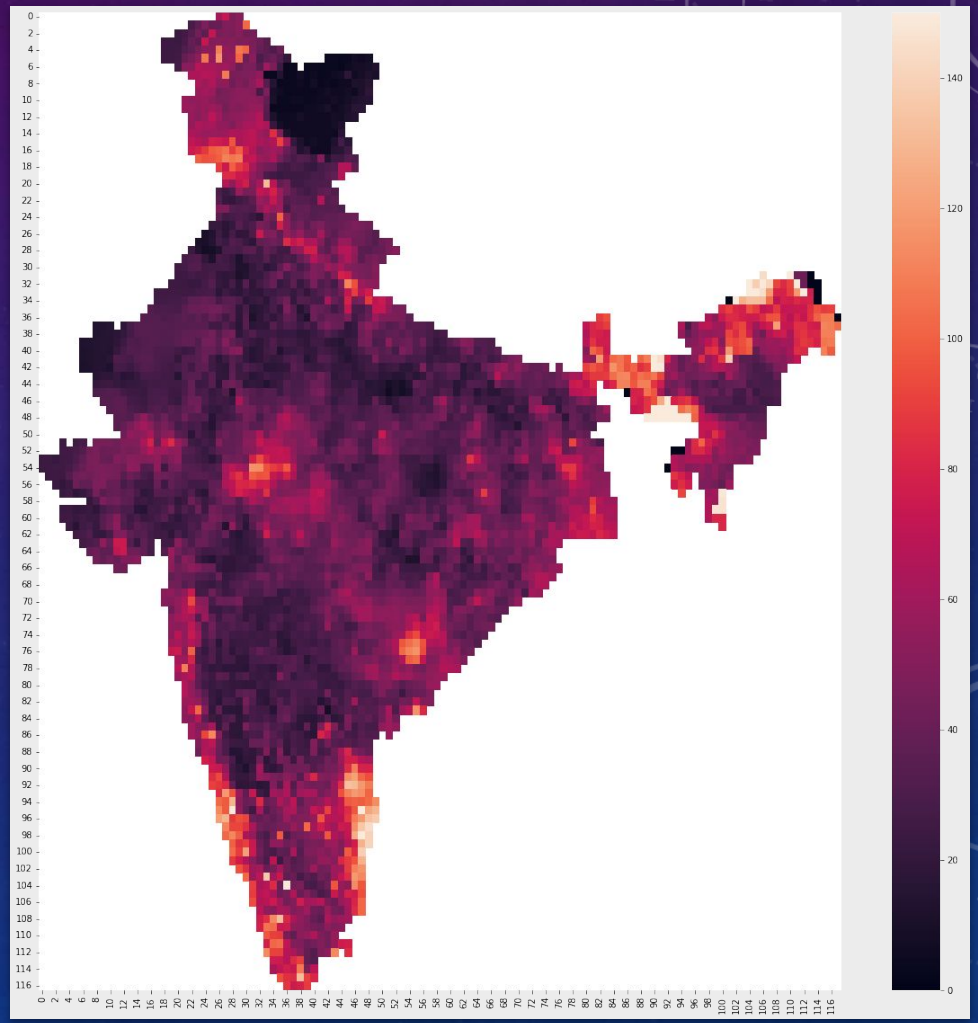
- The Unclipped Version of the Model gave output closer 0 for every input.
- The majority of the value lies in 0-600 range and the maximum value was 6881.4



MONSOON RAINFALL PREDICTION



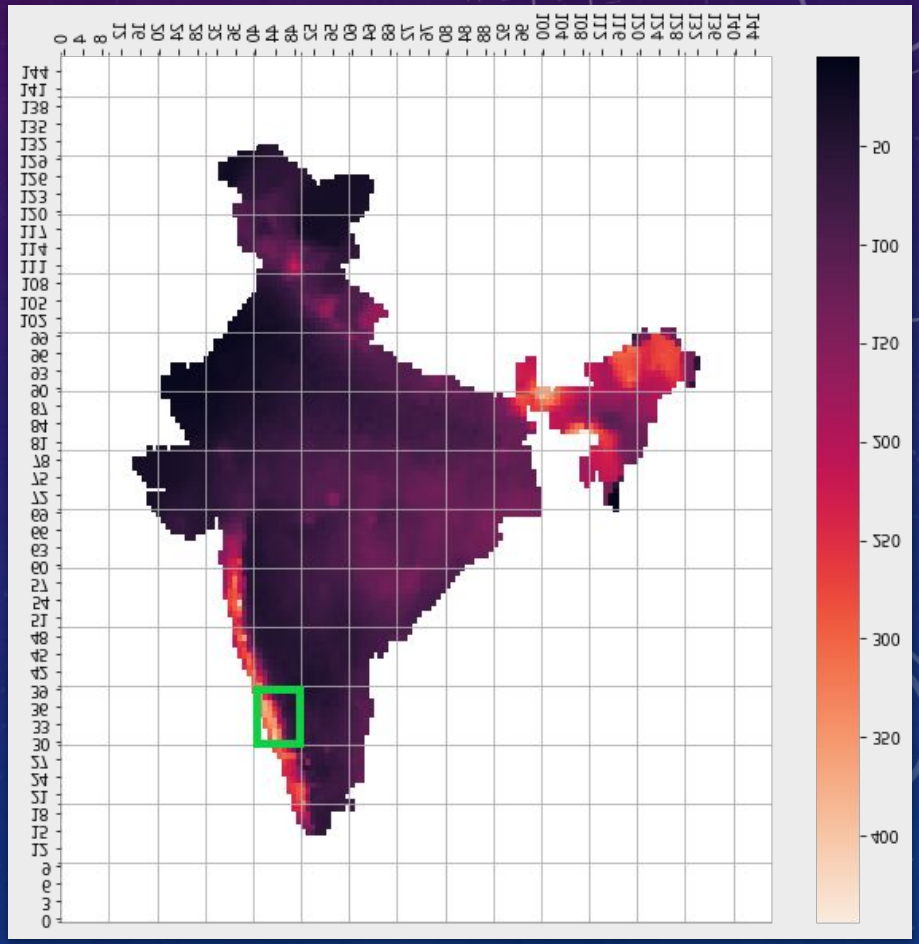
LSTM Monsoon Only



LSTM Full Month

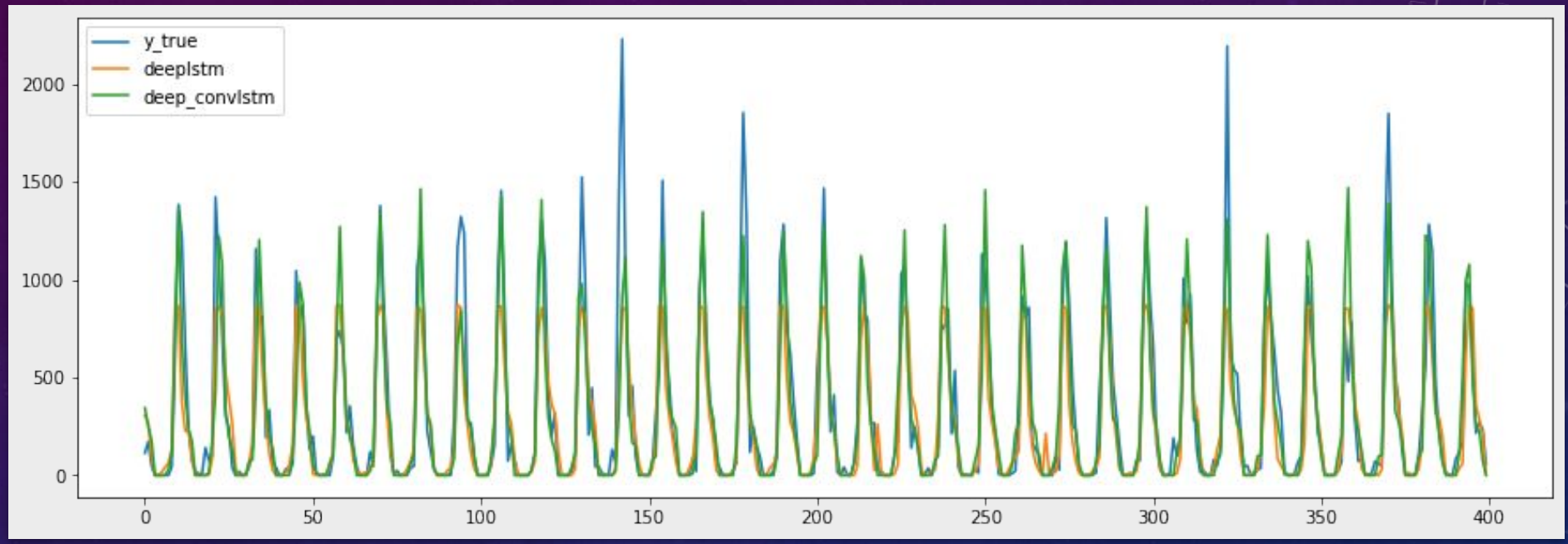
UPDATED CONV-LSTM MODEL

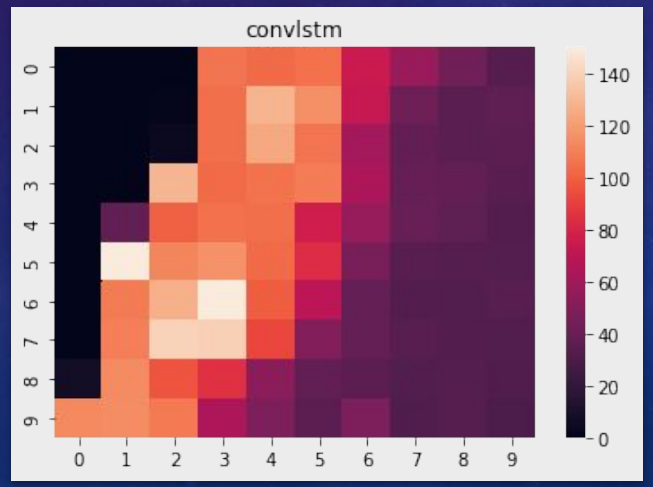
- Made to counter the previous issues of Conv-LSTM-based model
- Instead of predicting whole India, 10 x 10 grid blocks are predicted
- 28 x 28 grid blocks are used for spacial information, 12 horizon of time data
- Faster, works well with extreme values.
- Model for each block

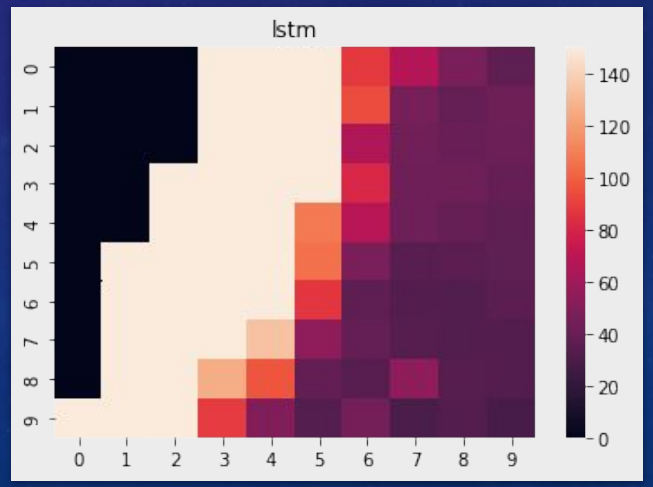


UPDATED CONV-LSTM V/S LSTM

Grid Point
277
13 , 75.25
(Lat,Long)







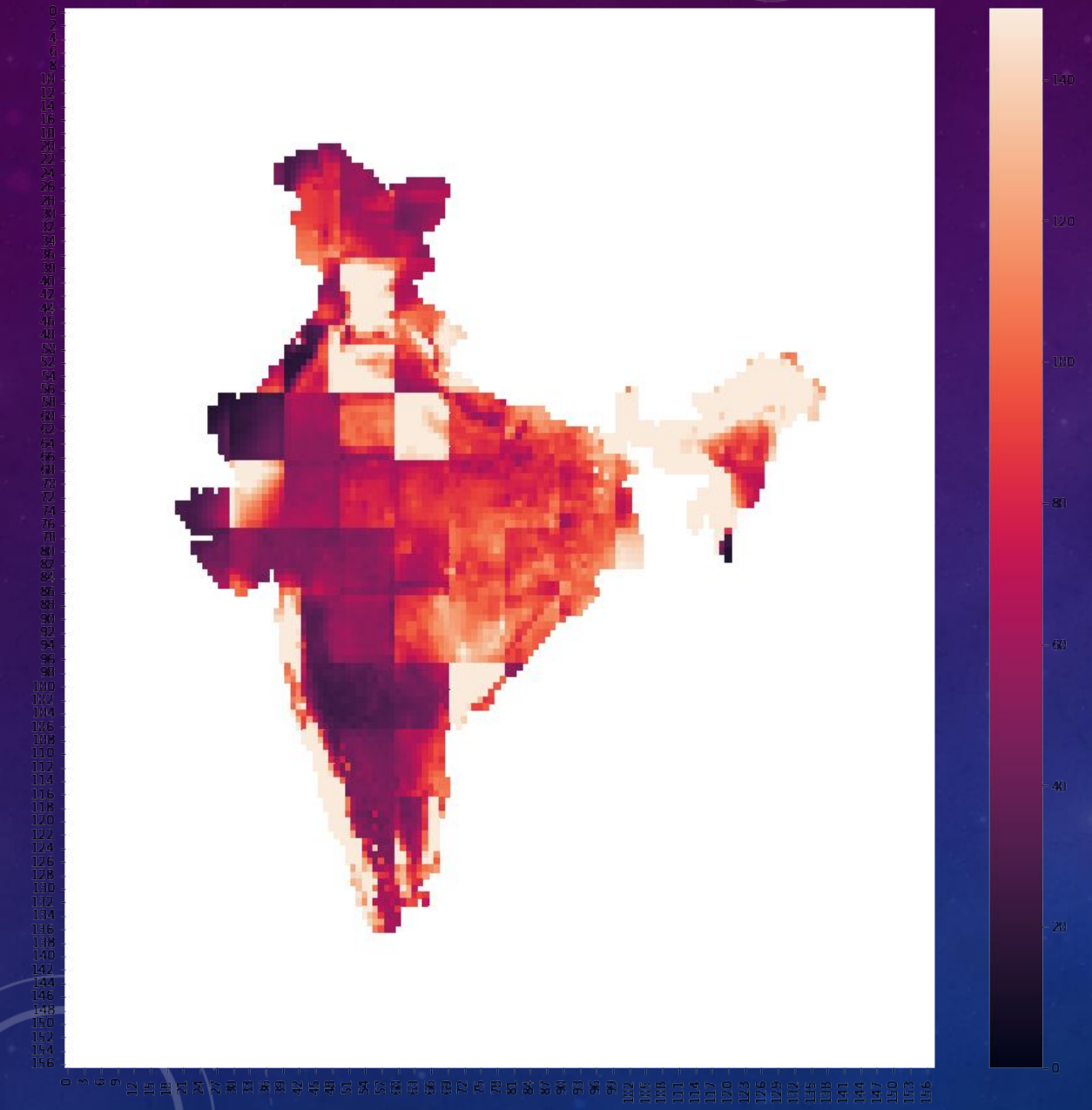
CONCLUSION AND FUTURE OBJECTIVES

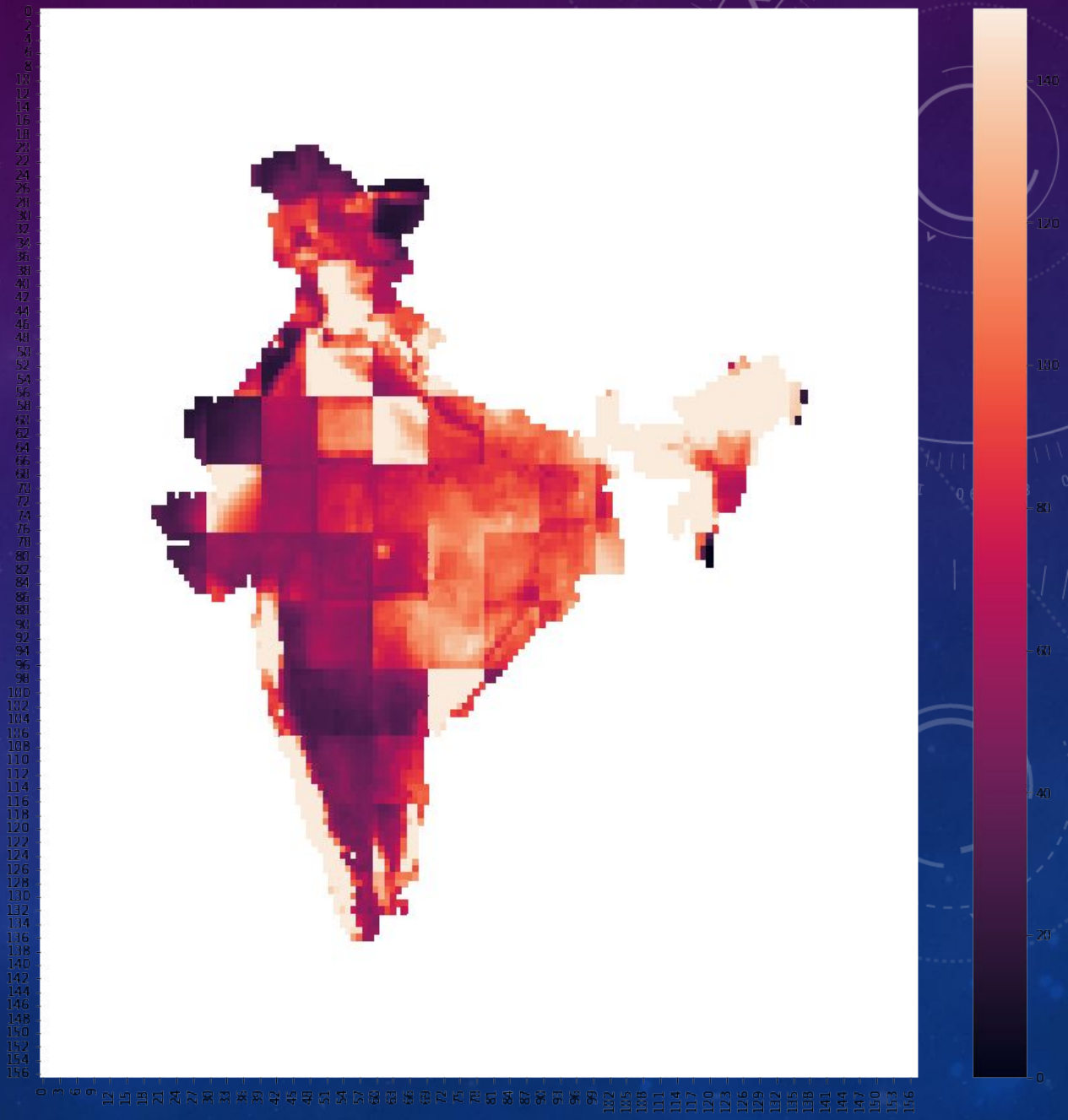
- Indian Monsoon Rainfall is harder to predict for each grid point with basic ML algorithms.
- The Extra Spacial Information improves quality of prediction.
- LSTM and AutoRegression (Time series) models were weak at predicting large values.

- Find a sweet spot for Block Dimension.
- Try out advanced Deep Learning Models (Graph Neural Networks).
- Connect more LSCIs.

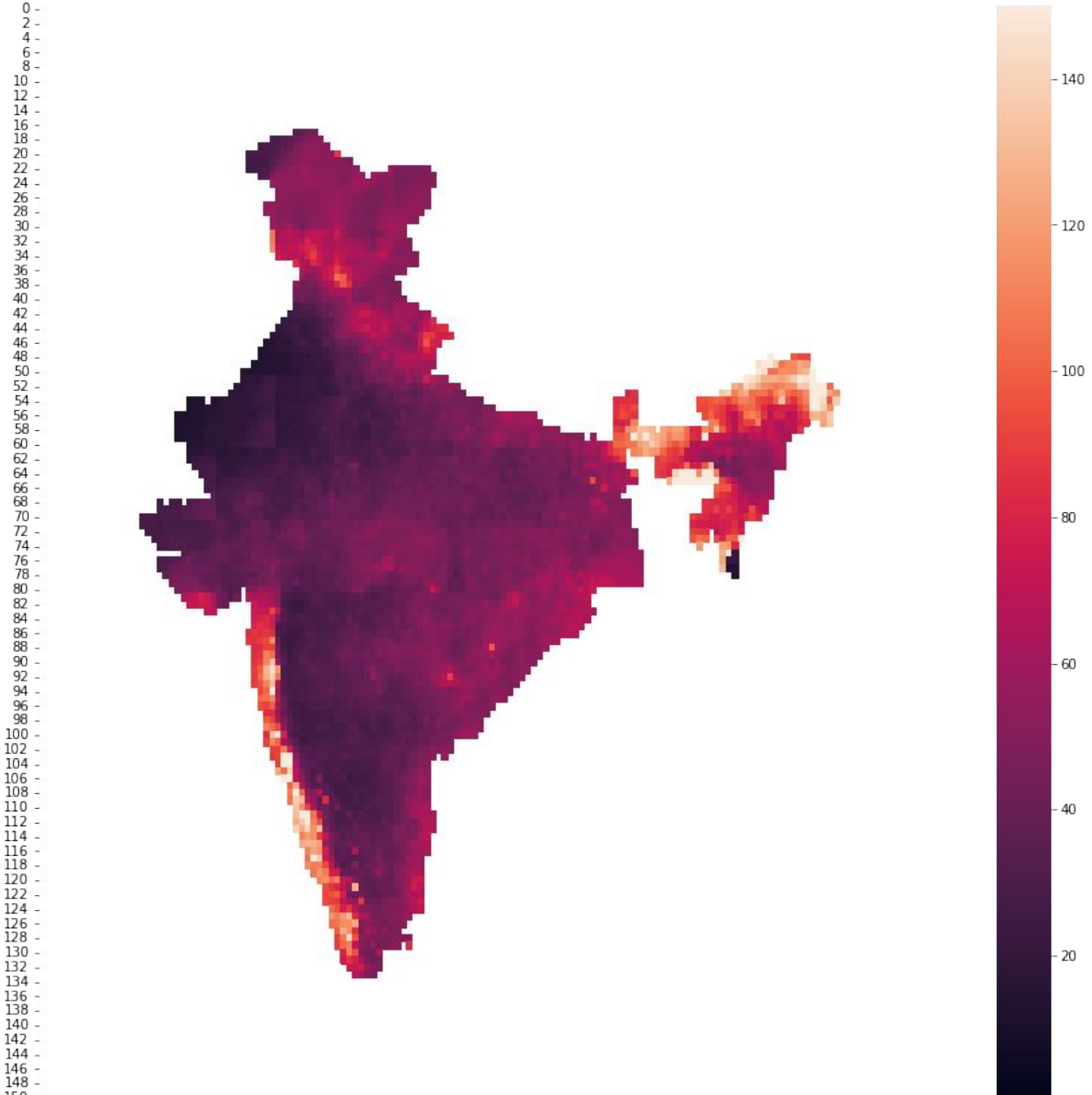
THANK YOU

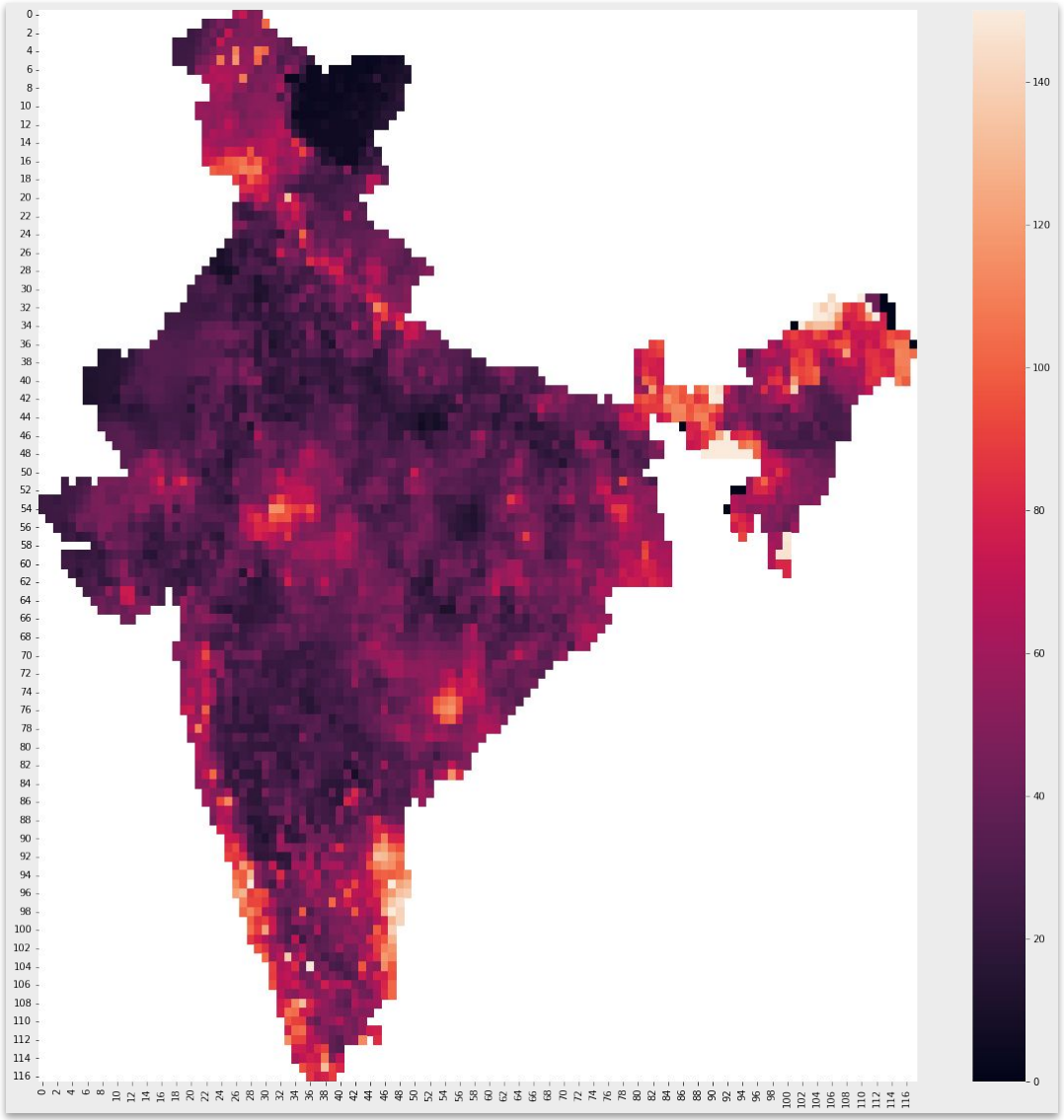
Thanks to my Mentors,
Dr. Sarmistha Singh
Dr. Chandra Shekar Lakshminarayanan


0
20
40
60
80
100
120
140

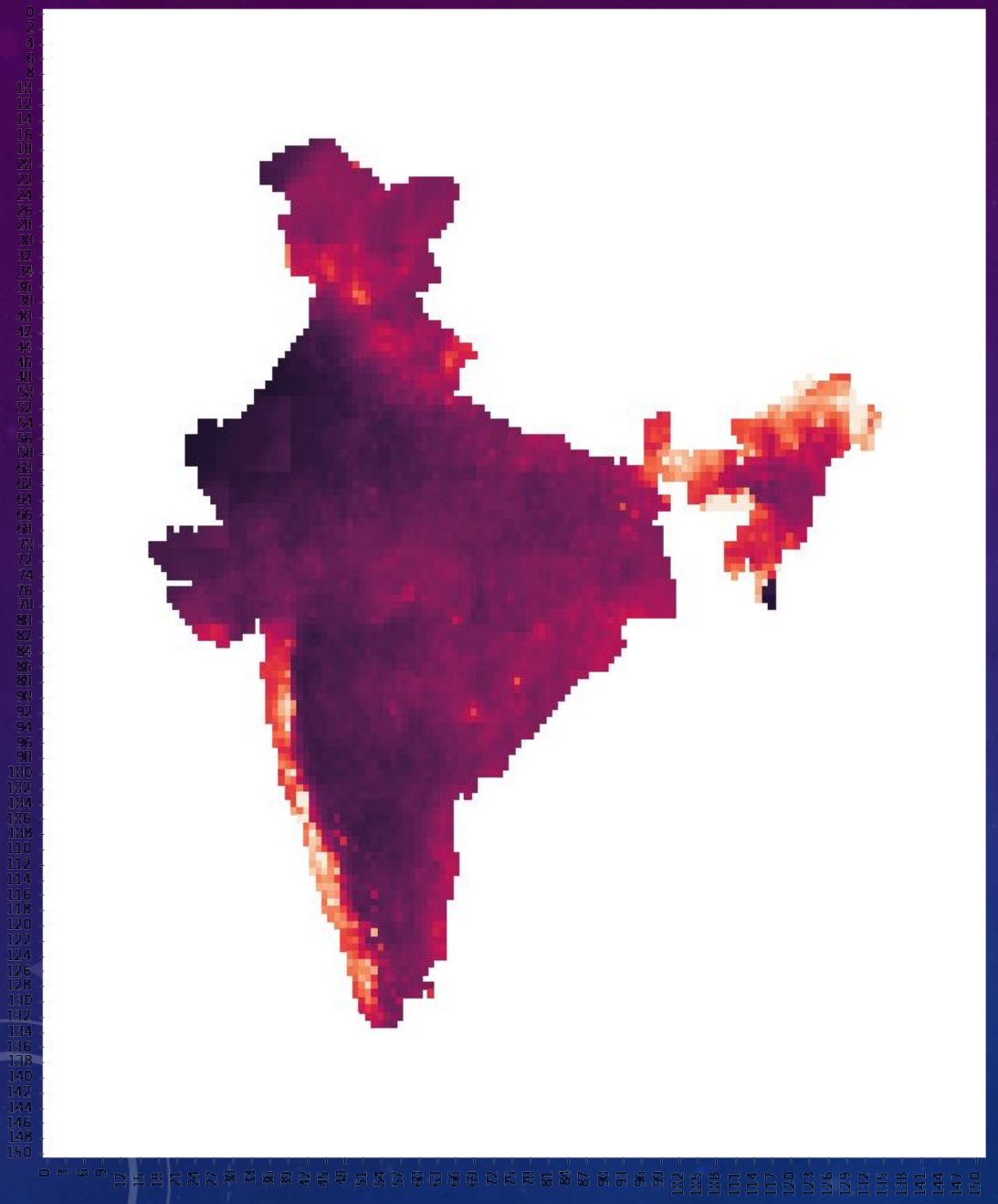

0
20
40
60
80
100
120
140

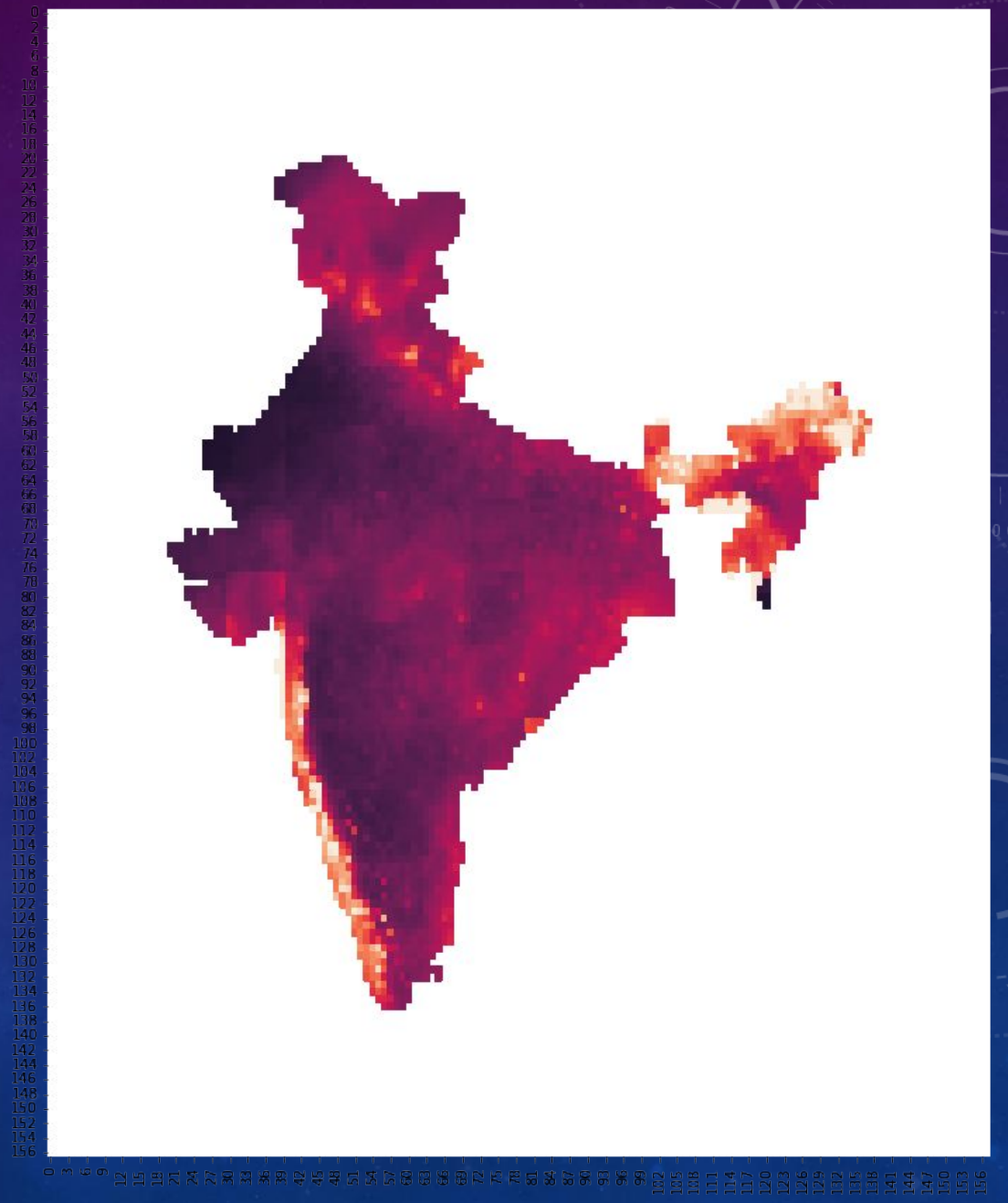
ConvLSTM new vs LSTM





With callbacks vs without callbacks





TODO

- Difference in error
- Range of rainfall