Use of Machine Learning Techniques for Seasonal and Subseasonal Studies and Predictions

.

Ravi S. Nanjundiah,

Indian Institute of Tropical Meteorology Pune,

Dr Sahai's Pathbreaking Work on AI

- One of the earliest work on Using ANN for Monsoon Prediction (2000)
- Prediction of SST anomalies using ANN (2006)
- Prediction of Active Break Cycles using Self Organizing Maps (2013)
- Use of Self Organizing Maps for ISO studies (2014)
- Bias Correction and Downscaling using Self Organizing Maps (2017)
- And Many more ….

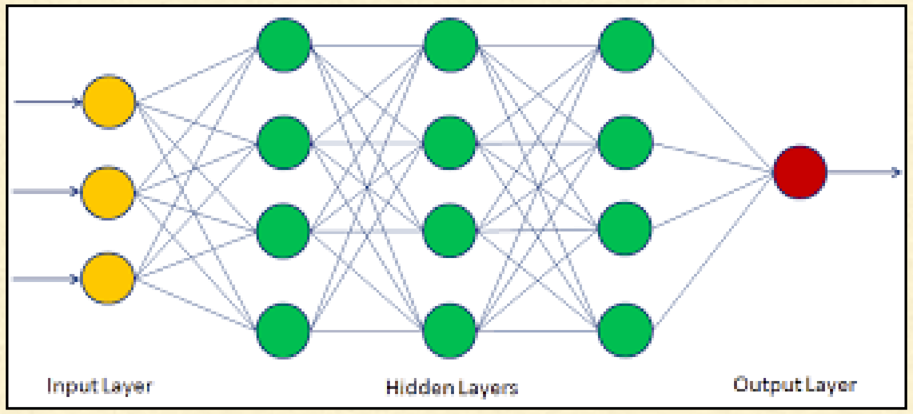
Outline

- Machine learning and motivation
- Exploring Climate Science with ML
    - Indian summer monsoon at seasonal (June-September) and sub-seasonal scales
    - ENSO and EQUINOO indices - remote impact on monsoons
    - Solar irradiance foroecasts - useful for Operating Solar Farms -extremly local
- Future directions of using ML in Climate Science
    - Weather at higher resolution and modelling at farm-level
    - Data Assimilation for weather & climate models
    - Bias correction for models outputs and forecasts
    - Prediction of Extremes
    - Hybridization between numerical and machine-learning models e.g. cloud convection

Machine Learning & Motivation

The Emergence of Machine Learning

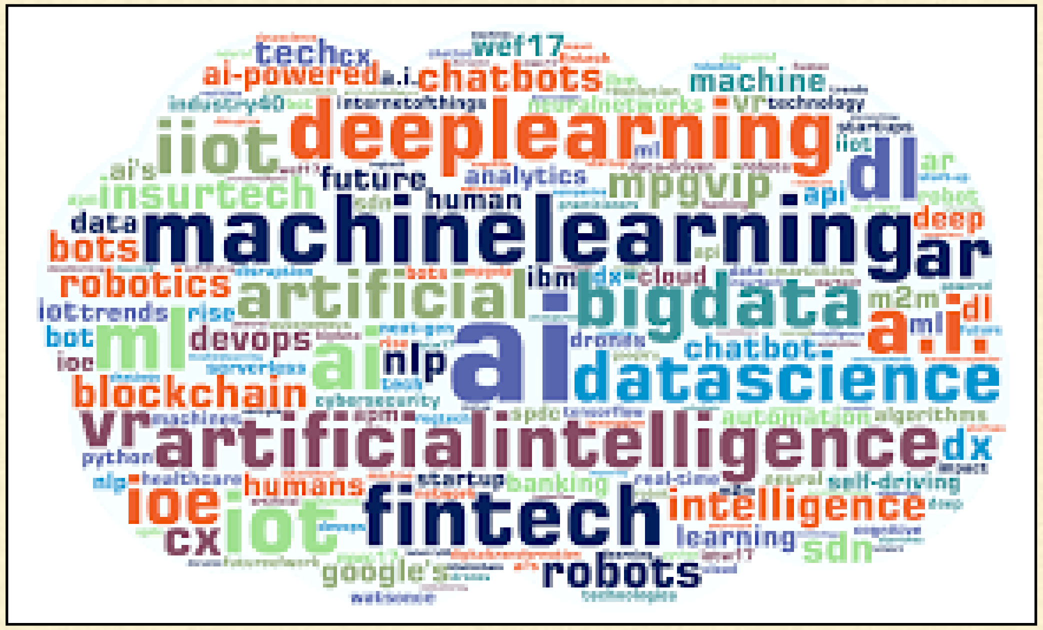
- Machine learning (ML) methods, and deep learning (DL), demonstrated impressive skills in reproducing complex spatiotemporal climatic processes



- The emergence of deep and machine learning is majorly

  due to:

  - the development of efficient and user-friendly libraries
  - the increasing computational capabilities (in particular the GPUs)
  - the access to large climatic datasets for training



Motivation

- Climate phenomena are complex events with high variability and uncertainty involved

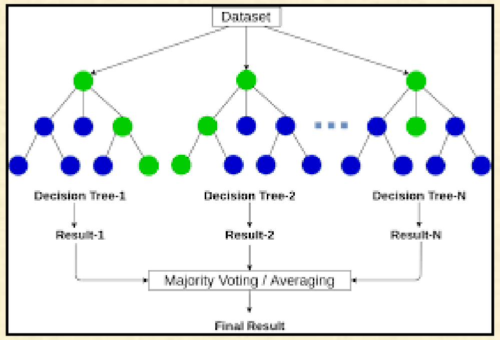
- The key motivation of using Machine Learning lies in the availability of huge climatic DATA…

- Data-driven artificial intelligence methods could help unravelling the intricacies of the phenomenon

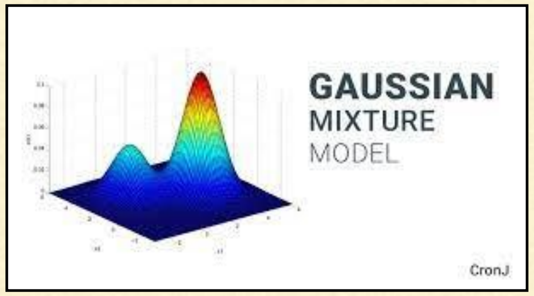
Machine Learning

Climate Science

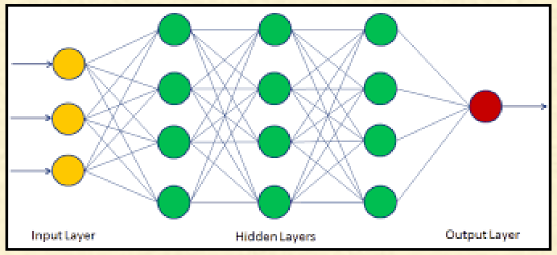
Machine Learning Models for Climate Science



- Forecasting of climate phenomenon
    - Framed as regression (numeric output) or classification (categorical output)
    - ML models: linear regression, logistic regression, decision tree, neural network

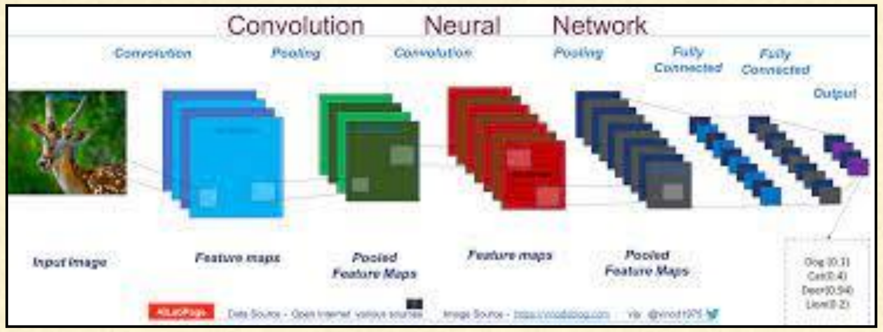


- Extreme events
    - Framed as anomaly detection problem
    - ML models: Gaussian mixture model, Bayesian change detection

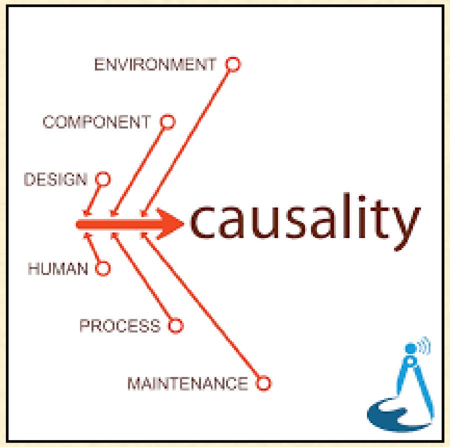


- Downscaling of climatic variables
    - Framed as mapping problem
    - ML models: artificial neural network, T-SNE neighbour embedding

Machine Learning Models for Climate Science

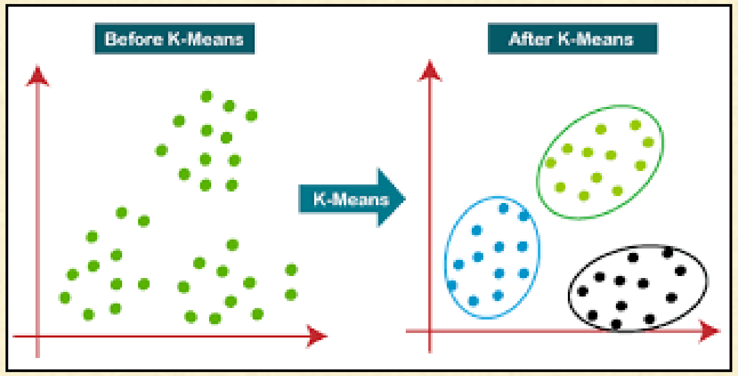


- Data assimilation
    - Framed as ensemble problem
    - ML models: convolutional neural network, ensemble regression model



- Study of climatic teleconnections
    - Framed as network analysis problem
    - ML models: causality method, community detection method

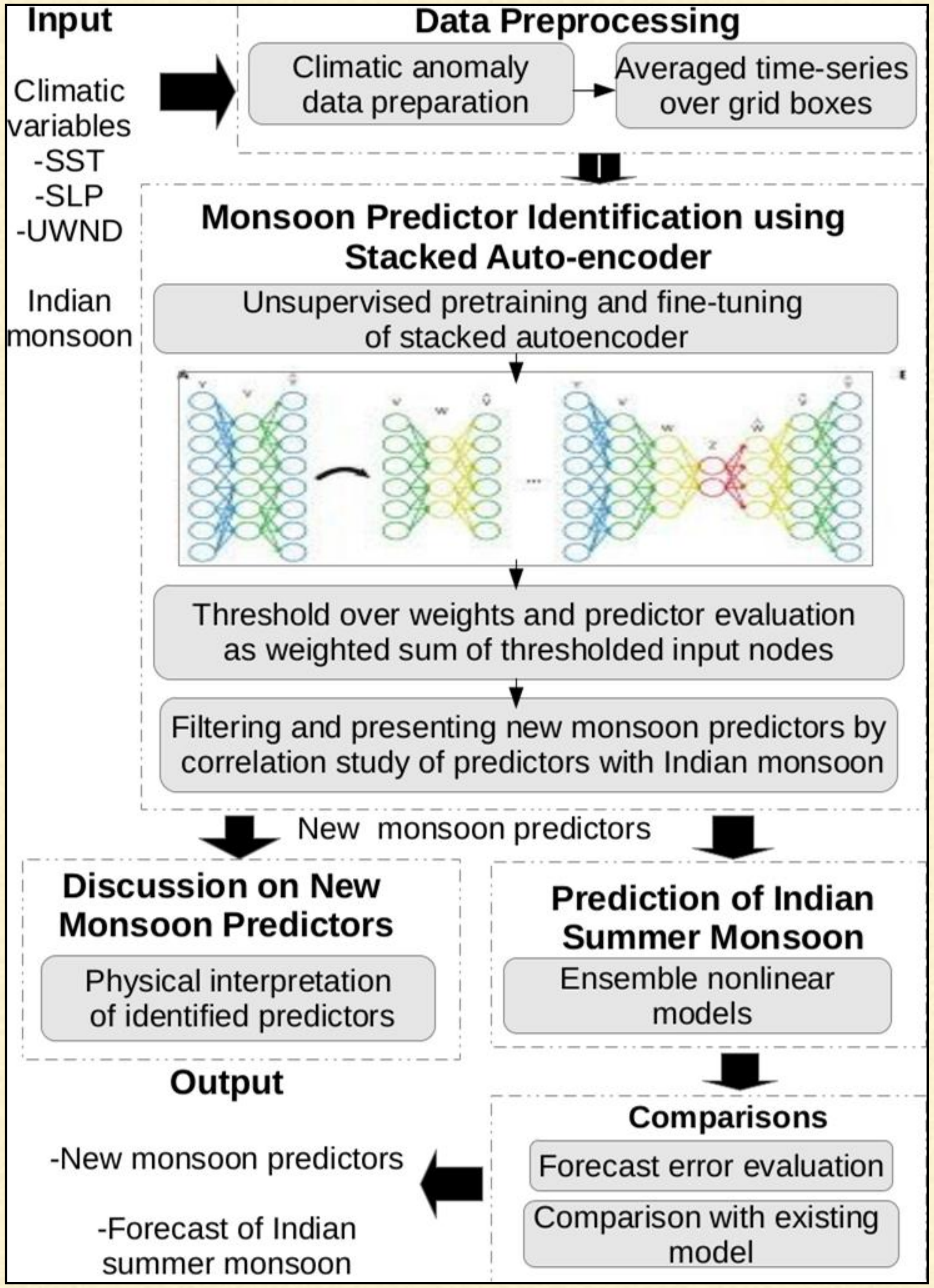
- Detection of climatic regimes
    - Framed as clustering problem
    - ML models: KNN clustering, DBSCAN clustering method



Exploring Climate Science with Machine Learning

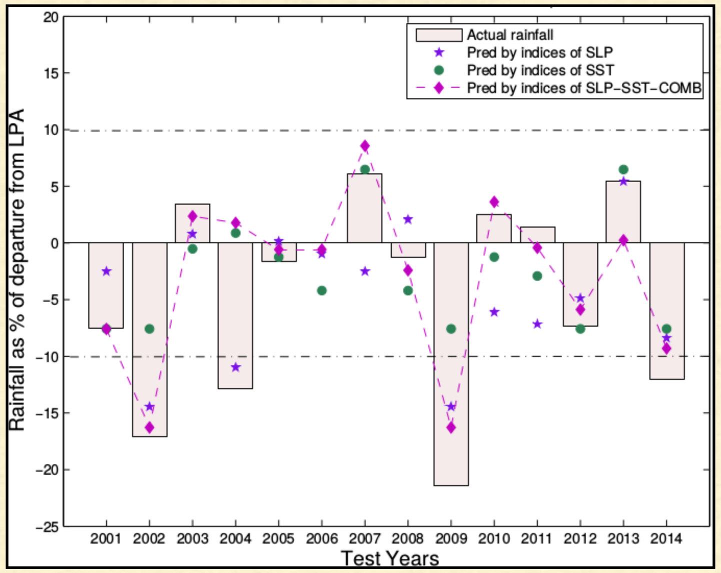
Stacked Autoencoder-Based Approach

- Automated feature learning and identifying new monsoon predictors
- Features are learnt at different abstraction at different levels
- Deeper the layer, more complex are the features
- Unsupervised feature learning, thresholding for feature extraction, supervised ranking



Stacked autoencoder approach

Prediction of All India Summer Monsoon

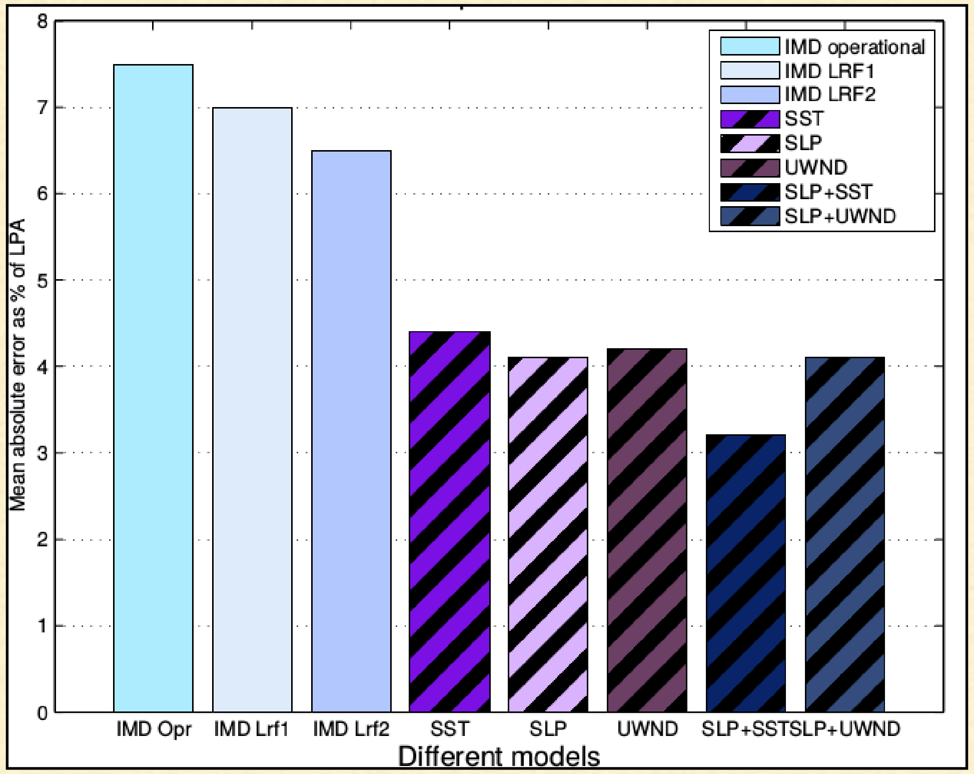


- Combined predictors of SLP+SST: 2.8%

  (with Predictors upto Mar i.e with about 3 months lead)

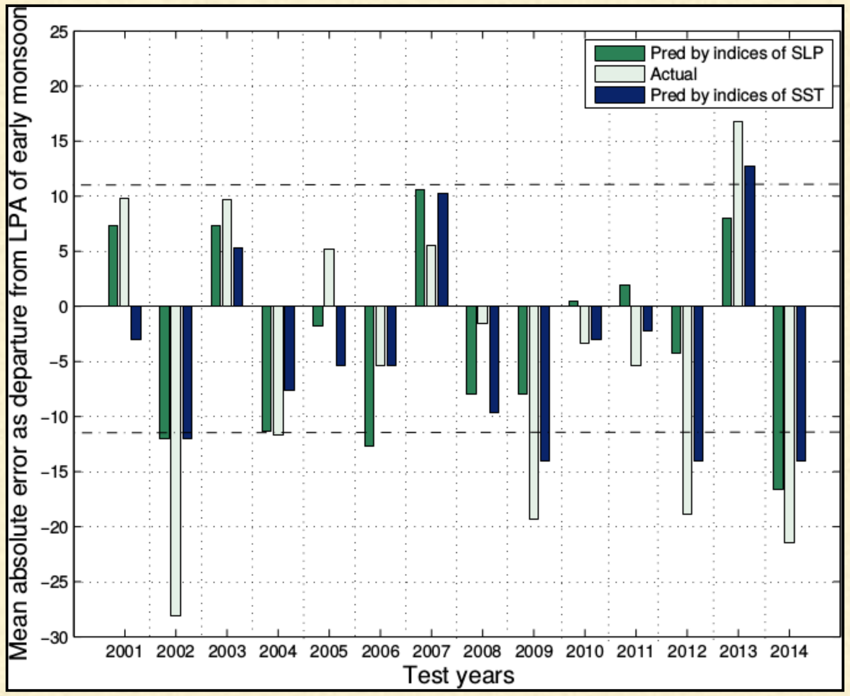
- IMD operational and PPR models give errors of 7.5%, 7.1%, and 6.5% in May, April, and June

- Stacked autoencoder based method performs superior to other statistical and numerical models



Prediction of Early-Late and Regional Monsoon

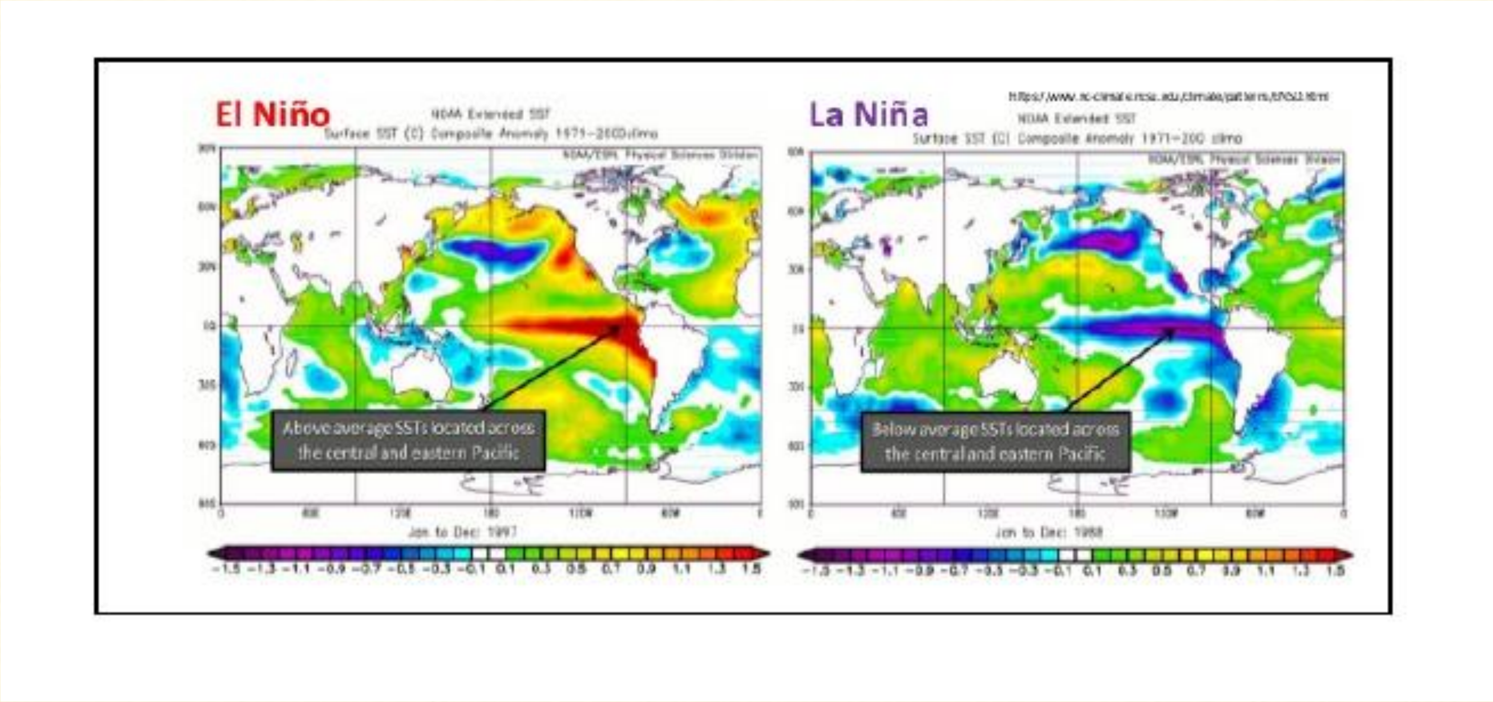
- It is important to know the behaviour of monsoons in its early and and later stages for various activities.
- Forecast early (June-July) and late monsoon (Aug-Sep) with 6.1% and 4.9% in April for early and late part of Monsoons

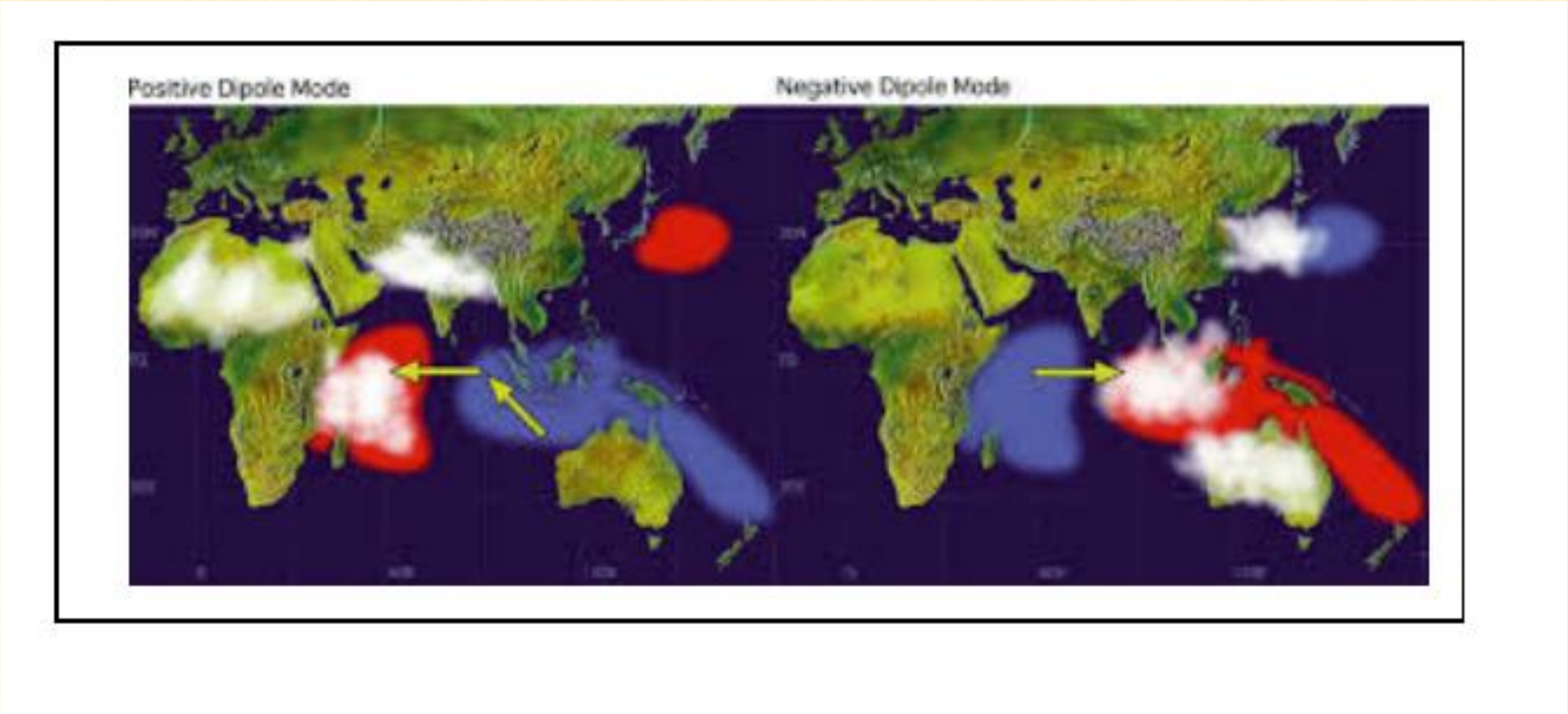


- We also need to have good estimates of monsoons over various parts of the country
- Appears to be better than IMD model for early/late and regional monsoon

Regional rainfall	**IMD model**	**Month**	**Proposed model**	**Month**
Central India	12.2	June	**4.8**	January
North-east India	7.8	June	**5.4**	March
North-west India	9.6	June	**6.1**	November
South-peninsular India	8.9	June	**5.3**	March

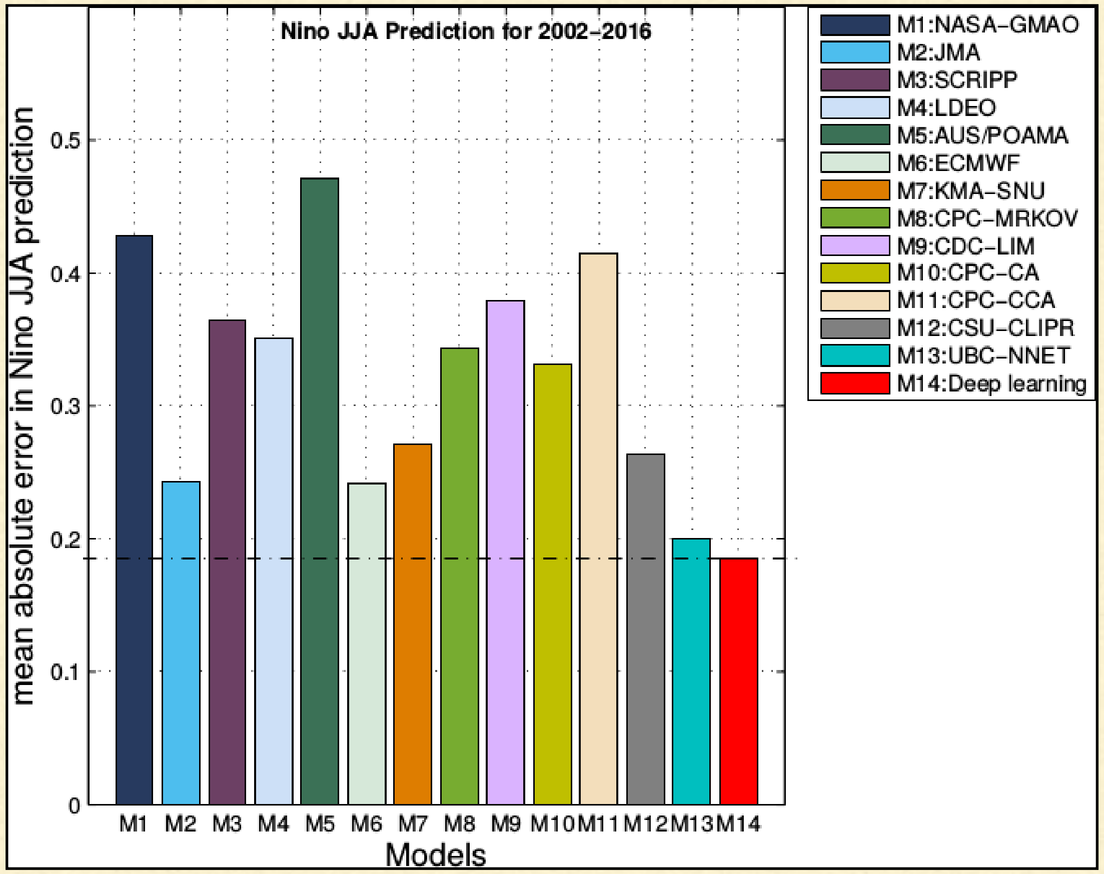
Prediction of EnSO and EQUINOO using Stacked Encoder

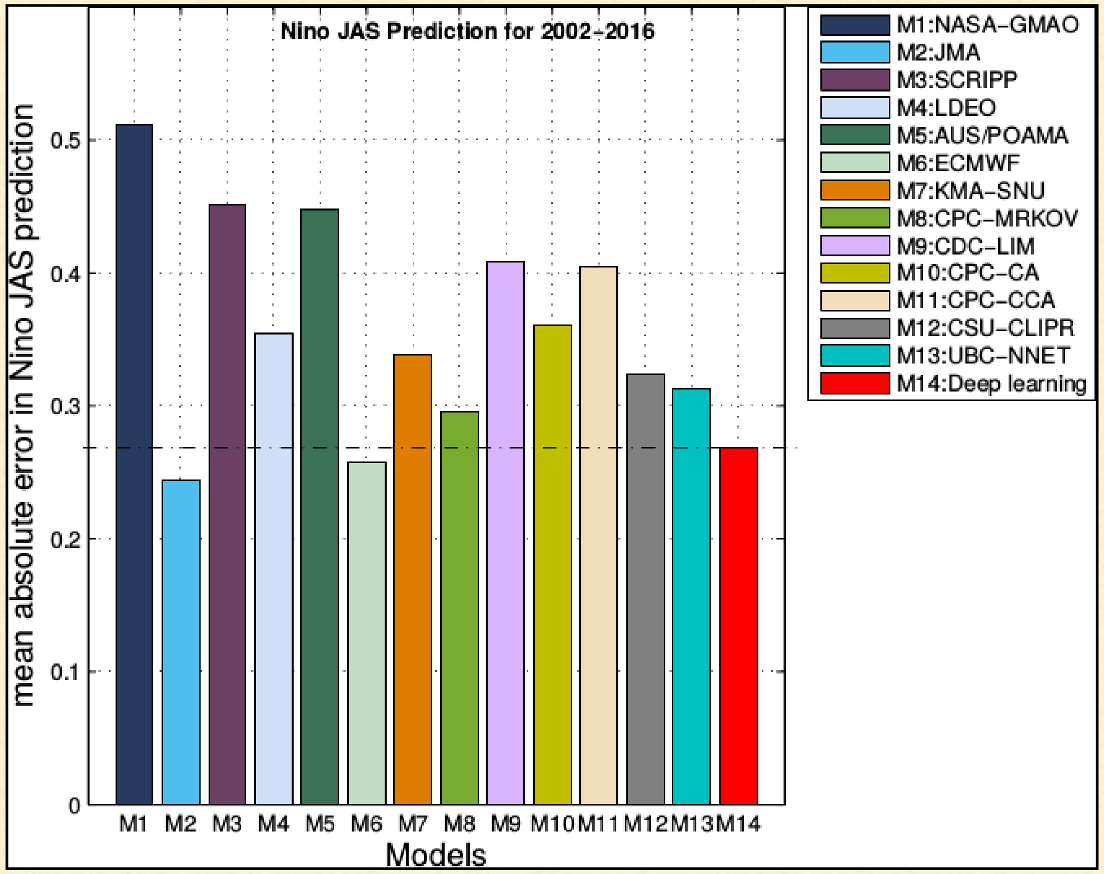




Prediction of ENSO and EQUINOO

- ENSO (El-Nino Southern Oscillation) is an ocean-atmospheric phenomena that affects global weather and climate
- EQUINOO (Equatorial Indian Ocean Oscillation) occurs over Indian Ocean
- Both effect the Indian Monsoon
- Predicted ENSO and EQUINOO with correlation coefficient of 0.87 and 0.88
- We attempted the EQUINOO prediction for the first time
- ML-based autoencoder model performs comparably (or better) than thirteen existing ENSO prediction models





- **Predicted ENSO for JJAS with correlation of 0.87**

Measures for ENSO	Values				
	JJAS	June	July	Aug.	Sep.
Correlation	0.87	0.88	0.88	0.84	0.87
Sensitivity	0.77	1.0	0.88	0.75	0.87
Specificity	0.85	1.0	1.0	0.87	0.87
Precision	0.87	1.0	1.0	0.85	0.87
Neg. pred. rate	0.75	1.0	0.87	0.77	0.87
Accuracy	81.2	100	93.7	81.2	87.5
F1 score	0.82	1.0	0.94	0.80	0.87

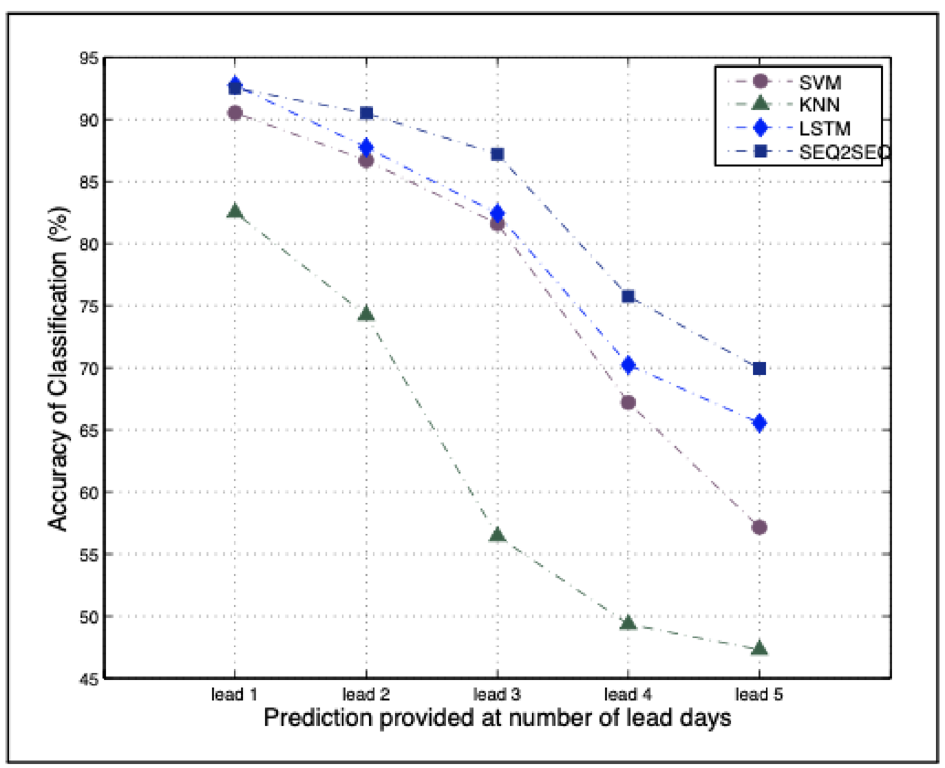
Prediction of EQUINOO Using Stacked Encoder

- EQUINOO prediction done at a lead of 6 months
- Perhaps the first such model

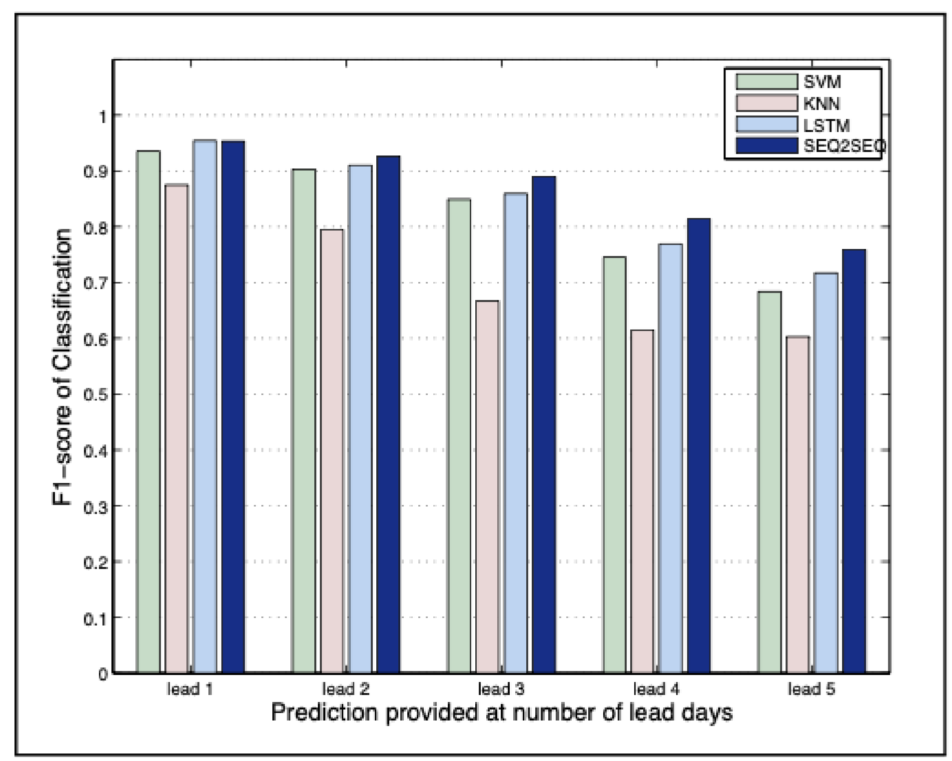
Measures for EQUINOO	Values				
	JJAS	June	July	Aug.	Sept.
Correlation	0.88	0.94	0.86	0.91	0.84
Sensitivity	1.0	0.85	0.80	1.0	0.90
Specificity	1.0	1.0	1.0	0.87	0.83
Precision	1.0	1.0	1.0	0.88	0.90
Neg. predictive rate	1.0	0.90	0.75	1.0	0.83
Accuracy	100	93.7	87.5	93.7	87.5
F1 score	1.0	0.92	0.88	0.94	0.90

Prediction of Monsoon at Sub-seasonal Scale

- LSTM and Sequence-to-Sequence models, capable of capturing long-distance temporal variation, used to predict monsoon at sub-seasonal scale
- At Subseasonal Scale, we are interested in cycles of active (rainy spells) and break spells (periods of low rainfall) within a season. Long breaks could result in droughts
- Convolutional neural network is also used which assists considering the spatial relationships between the climatic variables
- Performs superior than the traditional ML model



Accuracy of classification



F1-score of classification

Prediction of Active Break Spells Using CNN...

Comparing With SOM we find that CNN does better in Breaks and almost same skill scores for Normal and Active.

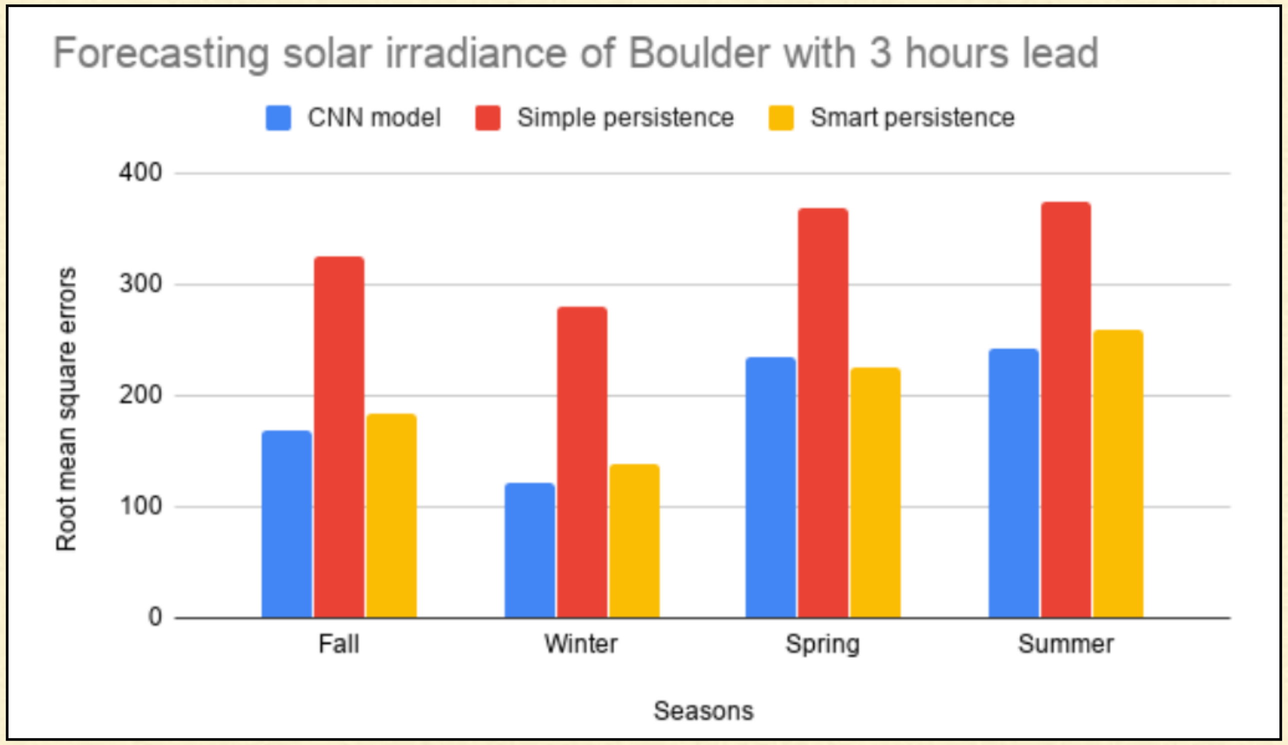
Evaluation metrics	Break		Normal		Active	
	CNN	SOM	CNN	SOM	CNN	SOM
Precision	0.74	0.48	0.58	0.56	0.58	0.56
Recall	0.62	0.49	0.74	0.60	0.50	0.46
F1- Score	0.66	0.48	0.64	0.57	0.53	0.50
AUC	0.84	0.74	0.69	0.58	0.81	0.78

Solar Irradiance Prediction: CNN-Based Method with Added Attention

- Prediction of Solar irradiance important for solar farm operations
- Convolutional neural networks (CNN) are capable of extracting features from data that have local spatial relations

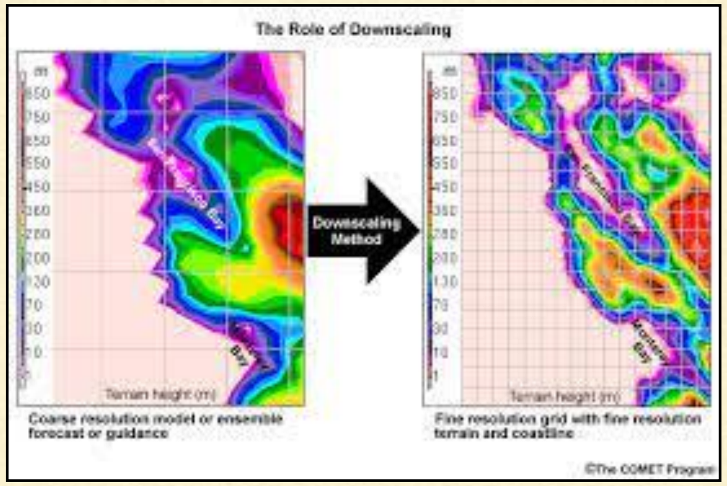
- We added dilation to the CNN kernel for capturing long-term dependencies

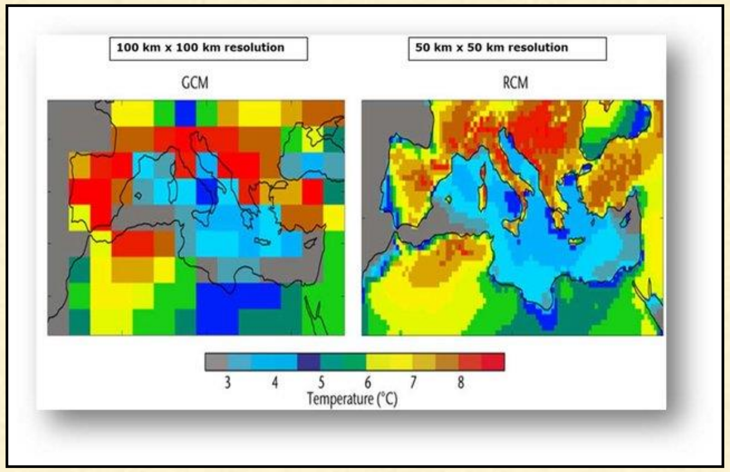
- Attention mechanism compels the model to focus on the parts of the input that bear a high impact

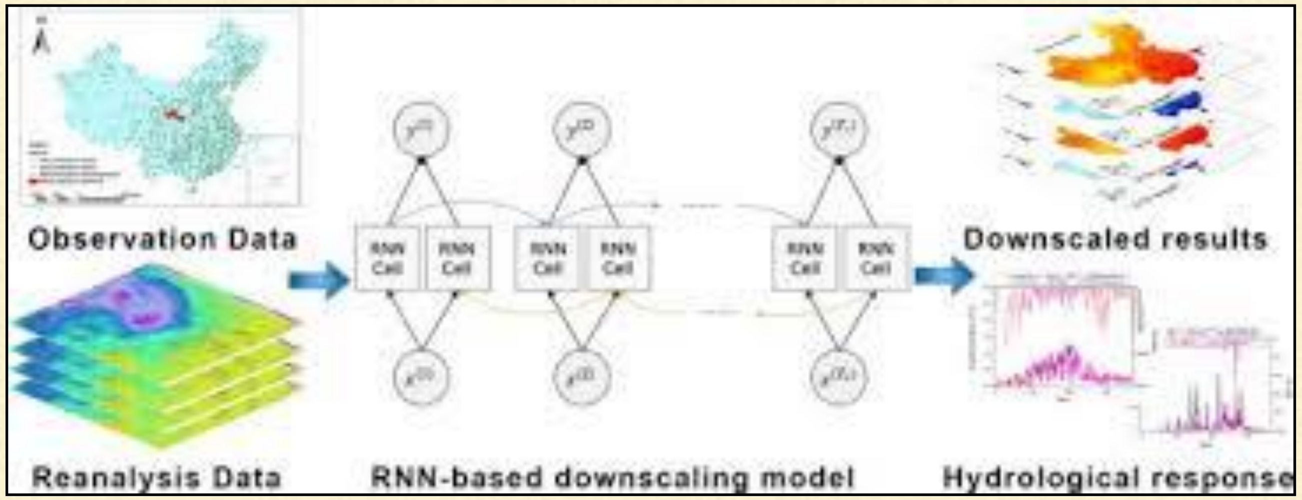


Downscaling

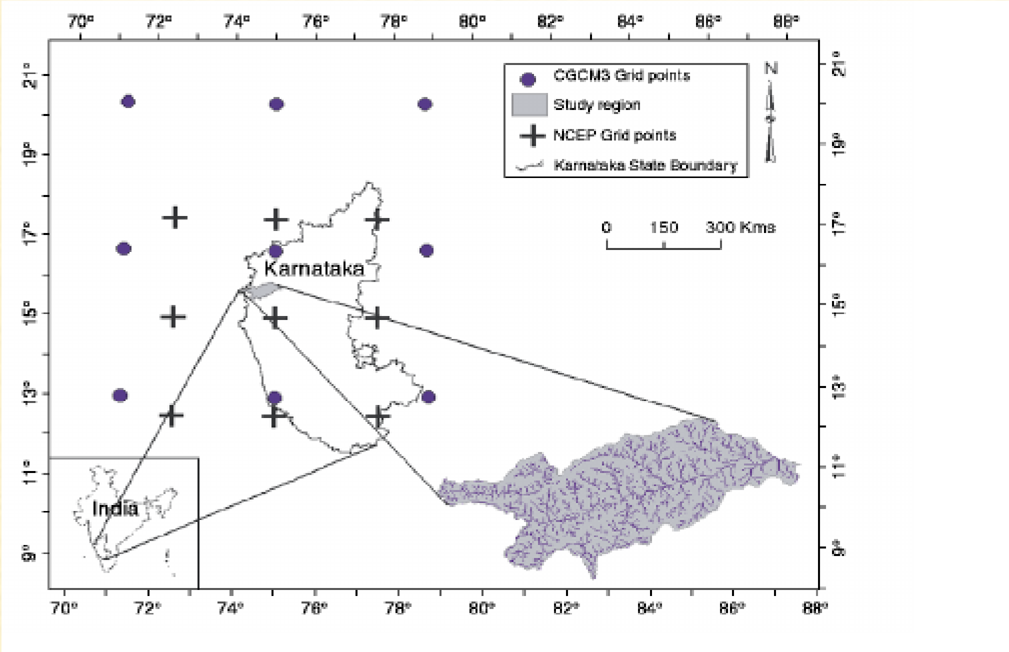
- Downscaling is the procedure of using large-scale climate models to provide climate predictions at finer temporal and spatial scales - very useful for stakeholders to provide information at finer scales for climate change scenario
- A variety of machine learning algorithm like artificial neural network, LSTM, multi-linear regression, support vector regressor could be used for downscaling

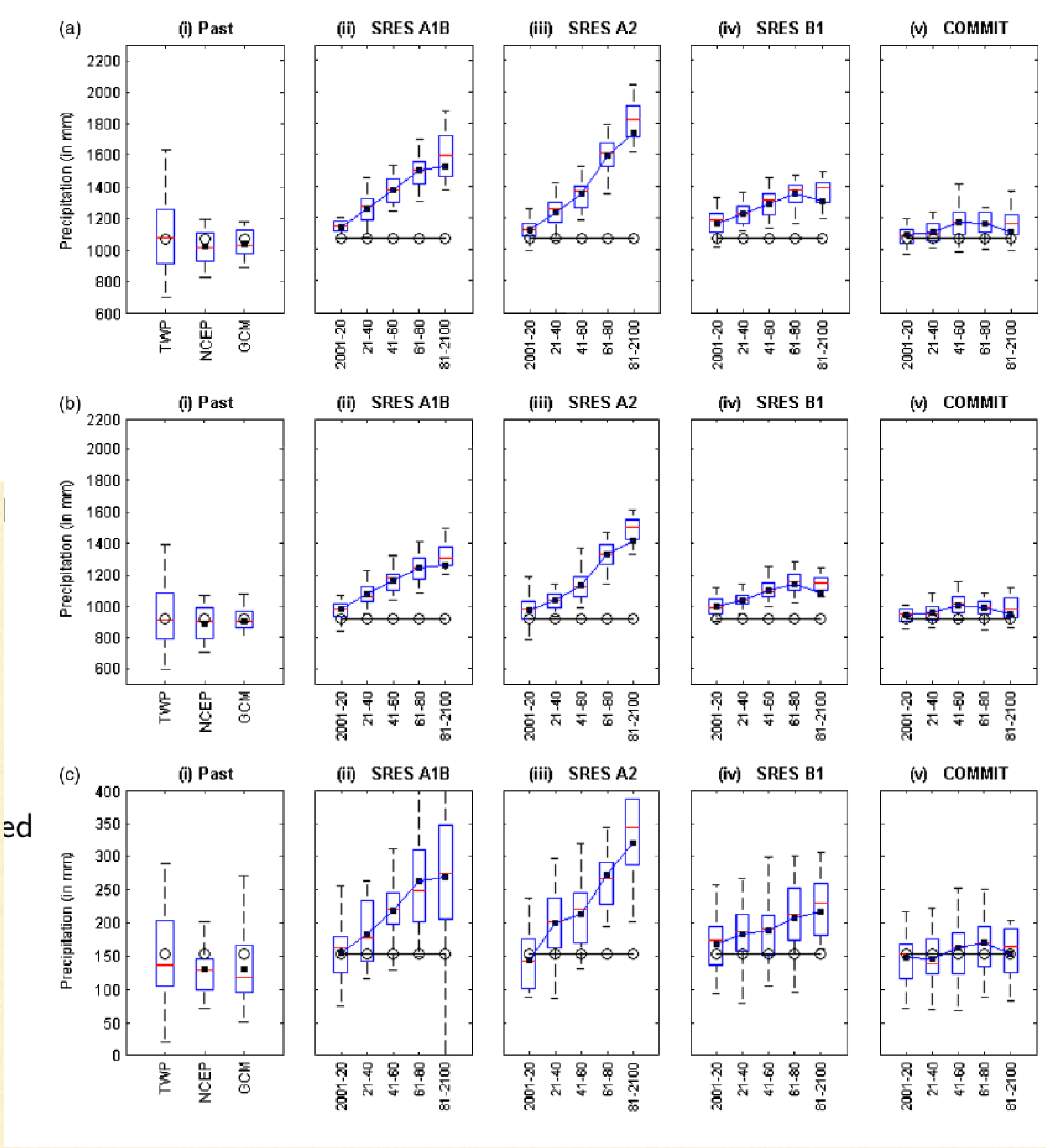






Downscaling



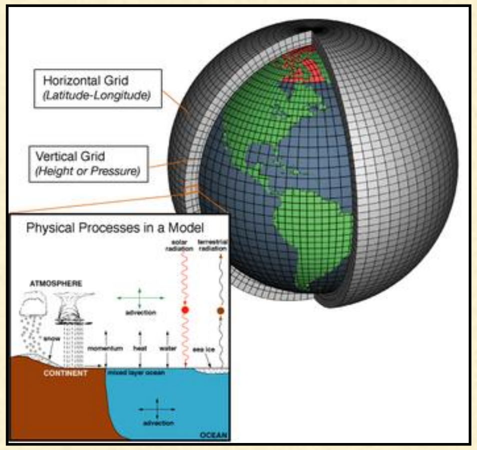


- ➔ Downscaling is creating information at finer resolutions from information at coarser scales
- ➔ Can be used to create information at finer scales from coarse resolltion simulations /forecasts
- ➔ Here we show an example of using SVM to downscale climate scenario simulations to that of a river basin in Northern Karnataka
- ➔ Using Climate change scenario we find that the rainfall could increase in the global warming scenario in the Malaprabha Basin

Future Directions

Weather at Higher Resolution

- We majorly focussed on climatic problems which are either small or medium scale



- More demanding problem is prediction or simulation at finer scales (in KMs)

- Challenges include both modelling (ML model and programming paradigm) and hardware support (configuration and capacity)

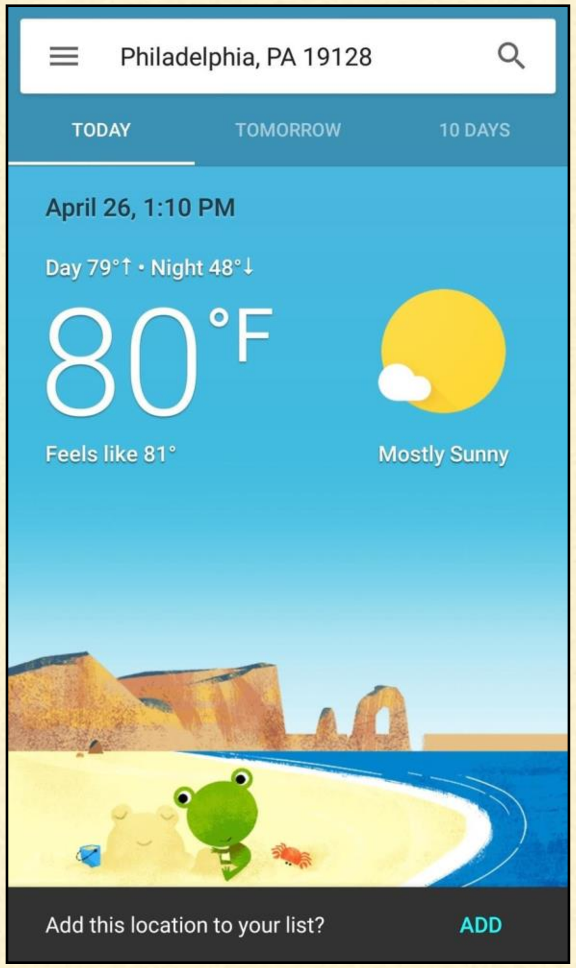
- We have to move to High Performance Computing for simulation and prediction at such higher scales

Weather Modelling at Farm-levels

- Google and IBM (The Weather Company) provides prediction of weather at much finer resolution in KMs
- Essentially produce forecasts at coarser scales, then using various Deep Learning Techniques and data at fine scales creates ultra-local forecasts.

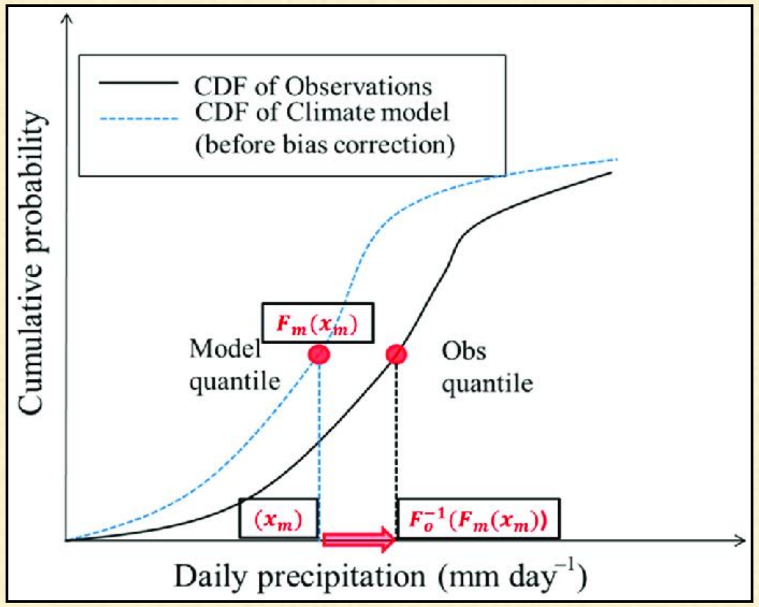
- Weather bench provides codes and comprehensive datasets to train the codes

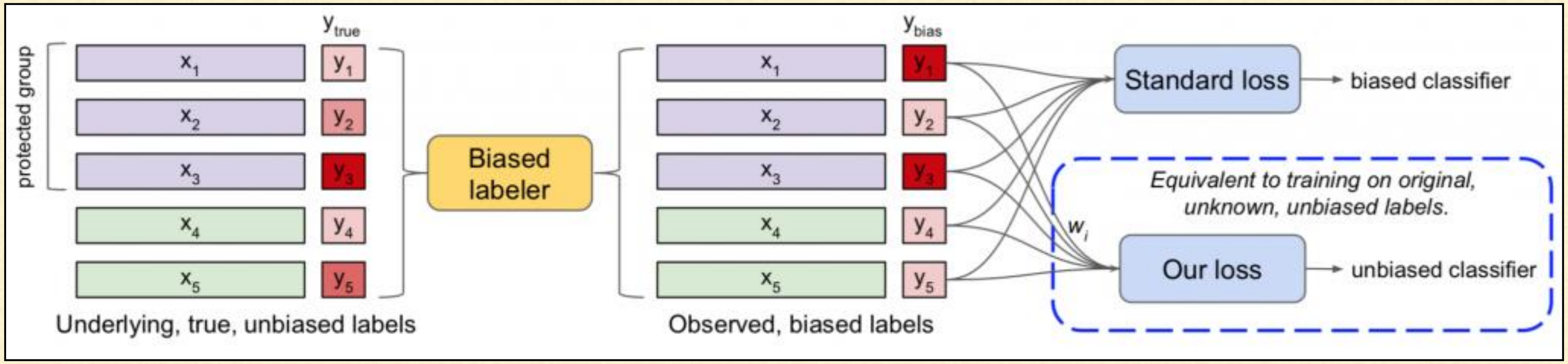
- One problem in modelling at such a finer scale is availability of ground-truth data -for specific farms with data available this work could be done -- needs interaction with industry and other stake holders



Bias Correction for Model Products

- Bias inherent in any action perception system (productive bias)
- Bias correction of numerical model outputs could be attempted with ML-based methods, like neural network or deep belief network
- Using observations and forecasts of past, Deep Learning models can be built to reduce model biases



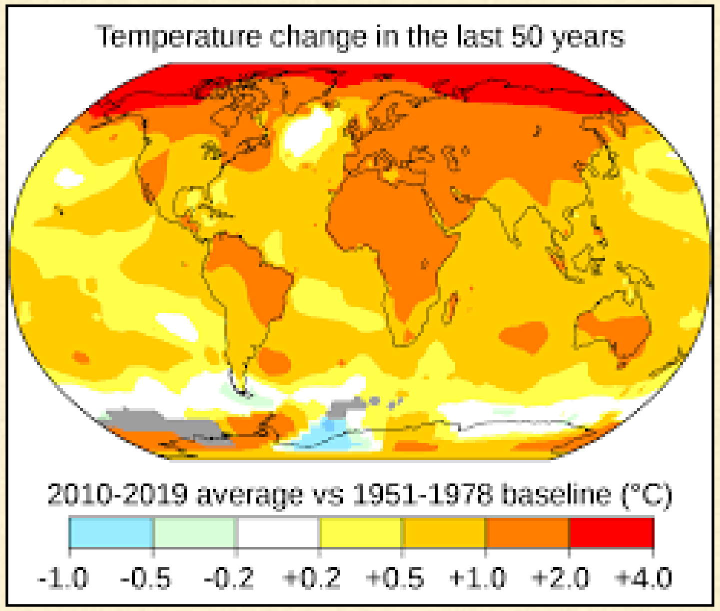


Extreme Events

- Extremes prediction should be considered under anomaly detection ML algorithm

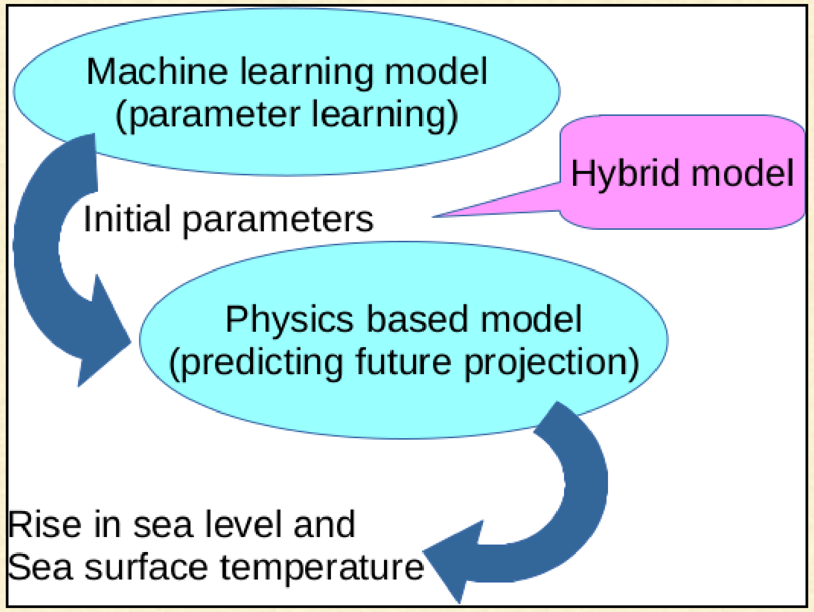
- Gaussian mixture model for identifying the precipitation extremes classes

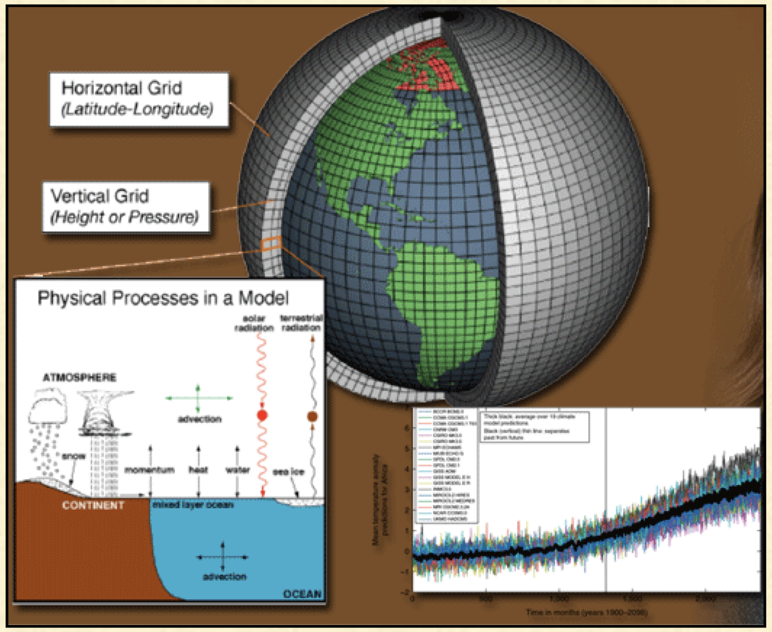
- Autoencoder-based reconstruction error could be monitored for detecting the extremes



Hybridization: Physics-based and Machine Learning

- Physics-based models: knowledge towards data– exploring how well current theory explains the data
- Machine Learning-based models: data towards knowledge– mathematical model describing relationships in data
- We cam combine the two develop hybrid models - use large data to develop empirical parts of the model
- Cumulus parameterisation - cloud scale convection is prescribed in models in empirical fashion. These models are built on small datasets
- Could be effectively replaced by Machine Learning methods embedded in numerical models.





Summary

- Machine Learning Appears to be an useful technique to study weather and climate
- We have explored various techniques to understand and predict climate and weather events
- Monsoon Prediction with AI/ML on various scales appears to have good skill
- Data Assimillation and Hybrid models could be the way forward to combine traditional Numerical Weather Prediction and Machine Learning Techniques
- With increased data availability from Satellites and other sources, better training of models for shorter scales can be attempted
- Downscaling can provide useful information to policy makers to decide about combating Climate Change
- Much Progress has been made but much remains to be done

Acknowledgement

Dr Moumita Saha (Currently with Phillips Innovation Centre, Bengaluru) for making me aware of the immense potential of AI/ML

Thank You !

ravisn@tropmet.res.in

Appendix

Stacked Autoencoder Based Identification of Monsoon Predictors for Aggregate, Early-Late and Regional Indian Monsoon

1. M. Saha, A. Santara, P. Mitra, A. Chakraborty, and R. S. Nanjundiah. “Pre- diction of the Indian Summer Monsoon Using Stacked Autoencoder and Ensemble Regression Model”, International Journal of Forecasting, 2020

2. M. Saha, P. Mitra, and R. S. Nanjundiah. “Deep Learning for Predicting Monsoon Rainfall over Homogeneous Regions of India”, Journal of Earth System Science, 2017

3. M. Saha, P. Mitra and R. S. Nanjundiah. “Autoencoder Based Climatic Index Discovery for Prediction of Indian Monsoon”, Meteorology and Atmospheric Physics, 2016

4. M. Saha, P. Mitra and R. S. Nanjundiah. “Predictor Discovery for Early-Late Indian Summer Monsoon Using Stacked Autoencoder”, International Conference on Computational Science (ICCS), 2016

Indian summer Monsoon

Indian summer monsoon

- Indian summer monsoon is a complex climatic phenomenon with uncertainty

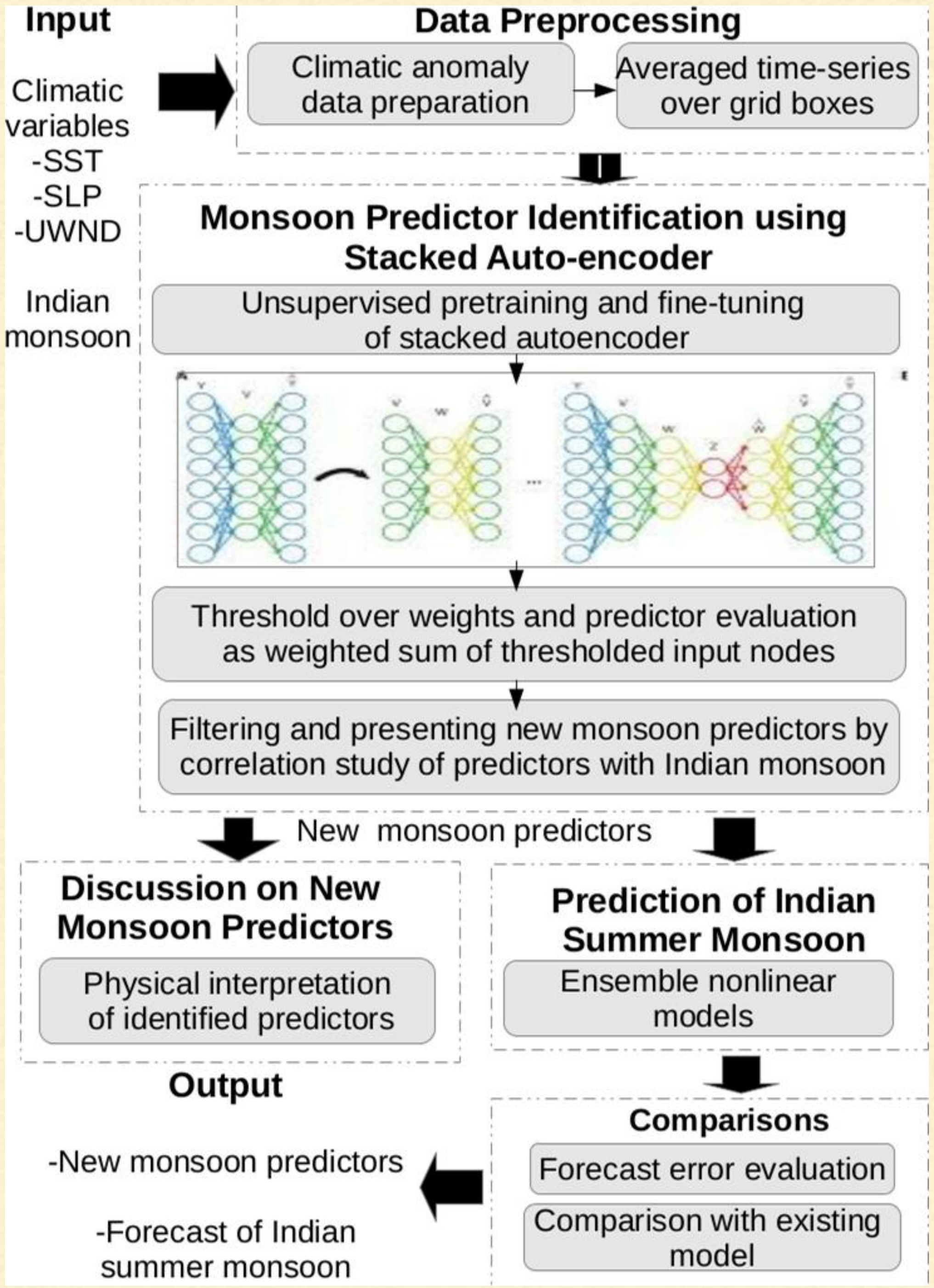
- Important to forecast Indian monsoon at temporal scale: early (June-July) and late (August-September) monsoons

- Spatial variation of monsoon is another important aspects: central, north-east, north-west and south-peninsular

- All four regions have different rainfall distribution and influencing factors

Stacked Autoencoder-Based Approach to Monsoon Prediction

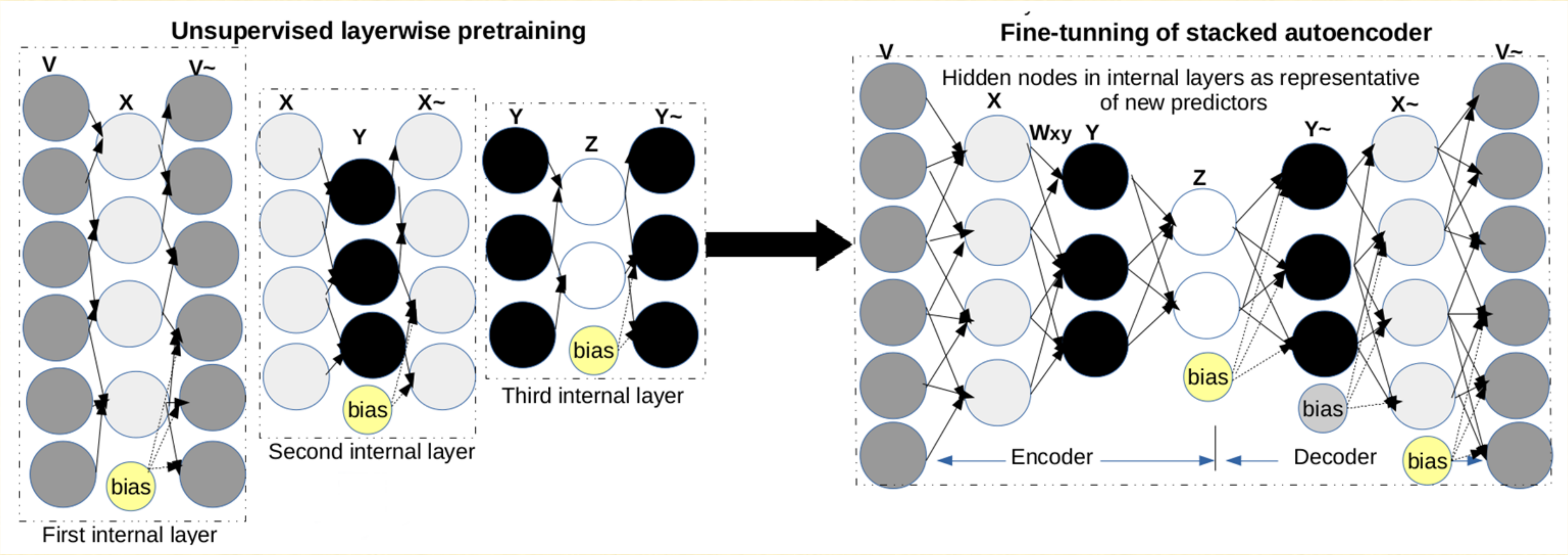
- Automated feature learning and identifying new monsoon predictors
- Features are learnt at different abstraction at different levels
- Deeper the layer, more complex are the features



Proposed approach with stacked autoencoder

Stacked Autoencoder

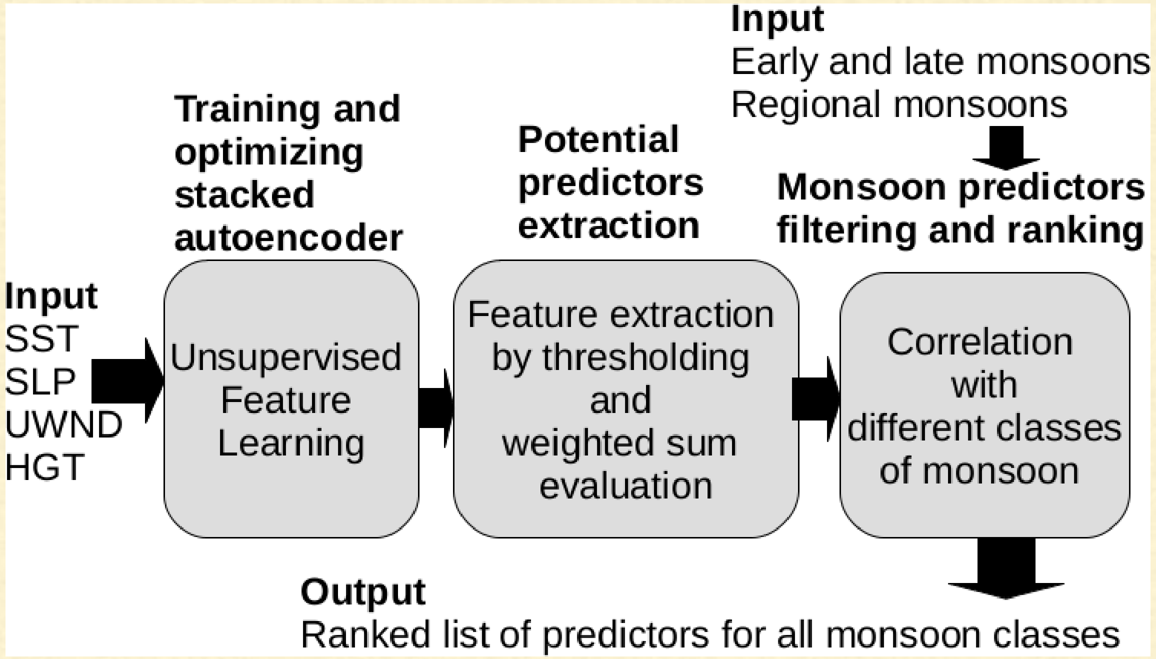
- Autoencoders are stacked to form deep network with output of previous autoencoder as input to current
- Unsupervised pre-training of one layer at a time
- Total network is fine-tuned using gradient descent algorithm



Stacked autoencoder

Identification of Monsoon Predictors

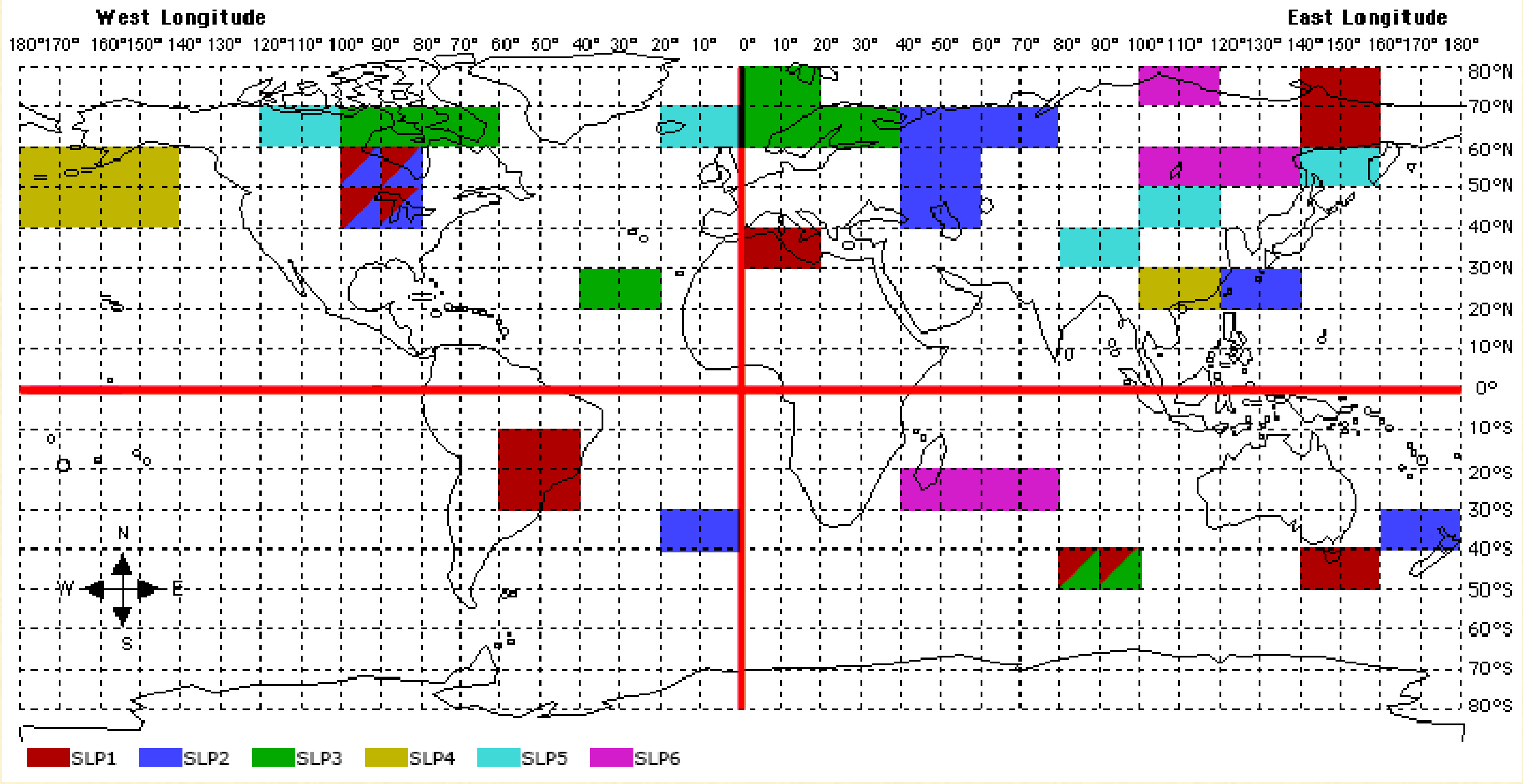
- Unsupervised feature learning: designed stacked autoencoder with grid variables as input
- Threshold for feature extraction: from internal layers: greater than twice standard deviation from mean
- Supervised ranking: based on their correlation with different categories of monsoons



Three-step method

Prediction Model with Identified Predictors

- Different regions are combined non-linearly as stacked autoencoder uses tan-hyperbolic
- Predictors as weighted sum of geographically distant regions



Identified SLP predictors for aggregate Indian monsoon

Prediction Accuracy for Aggregate Indian Monsoon

- SST, SLP, and UWND show errors of 4.3%, 4.0%, and 4.1% in predicting aggregate monsoon in April

- Accuracy in prediction increased with more composite features at the deeper layers

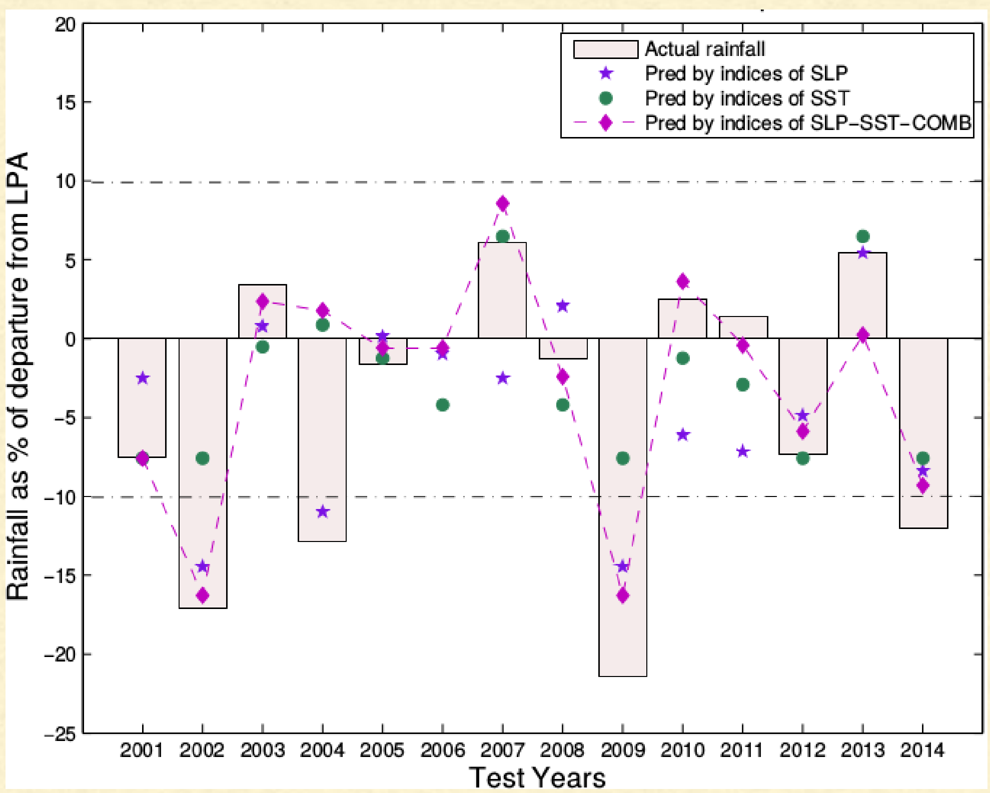
Table: Absolute mean errors(%) for 2001–2014

	Identified SST predictors of first layer				
	D1	**D2**	**D3**	**D4**	**D5**
RegTree	6.6	6.4	5.9	6.2	**5.4**
	Identified SST predictors of second layer				
RegTree	5.5	**4.4**	5.1	5.3	5.1
	Identified SST predictors of third layer				
RegTree	**4.3**	4.8	4.9	-	-

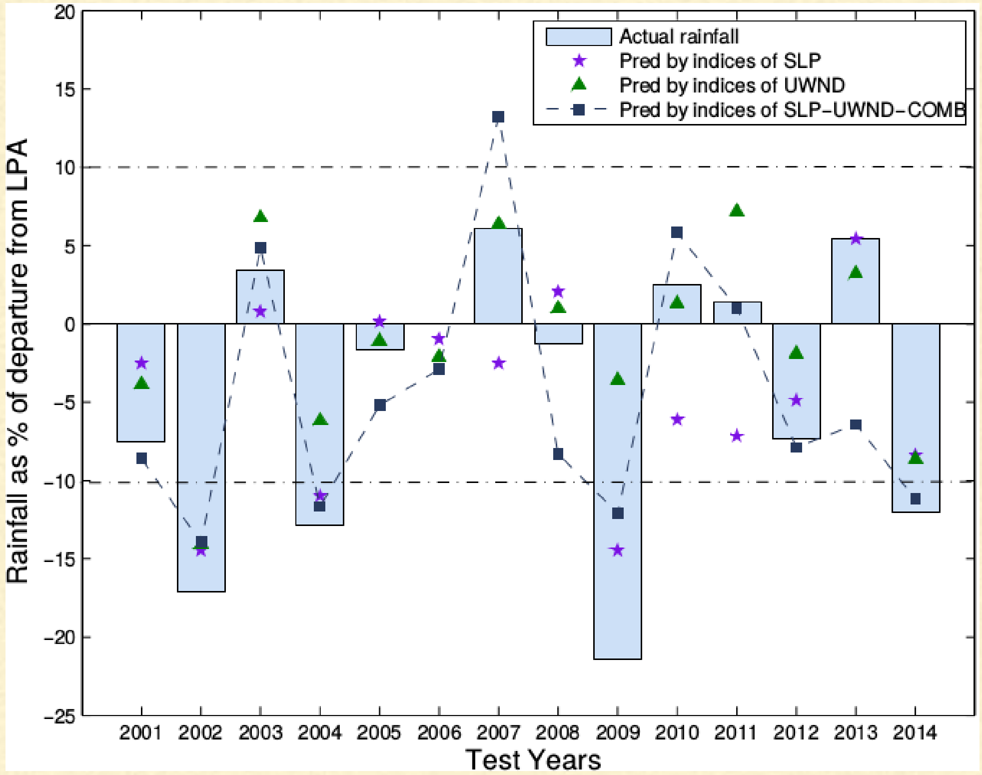
Performance of Combined Predictors over Individual

- Combined predictors of SLP+SST: 2.8%; Individual SLP: 4.0% and Individual SST: 4.3%

- Combined predictors of SLP+UWND: 3.7%; Individual SLP: 4.0% and Individual UWND: 4.1%



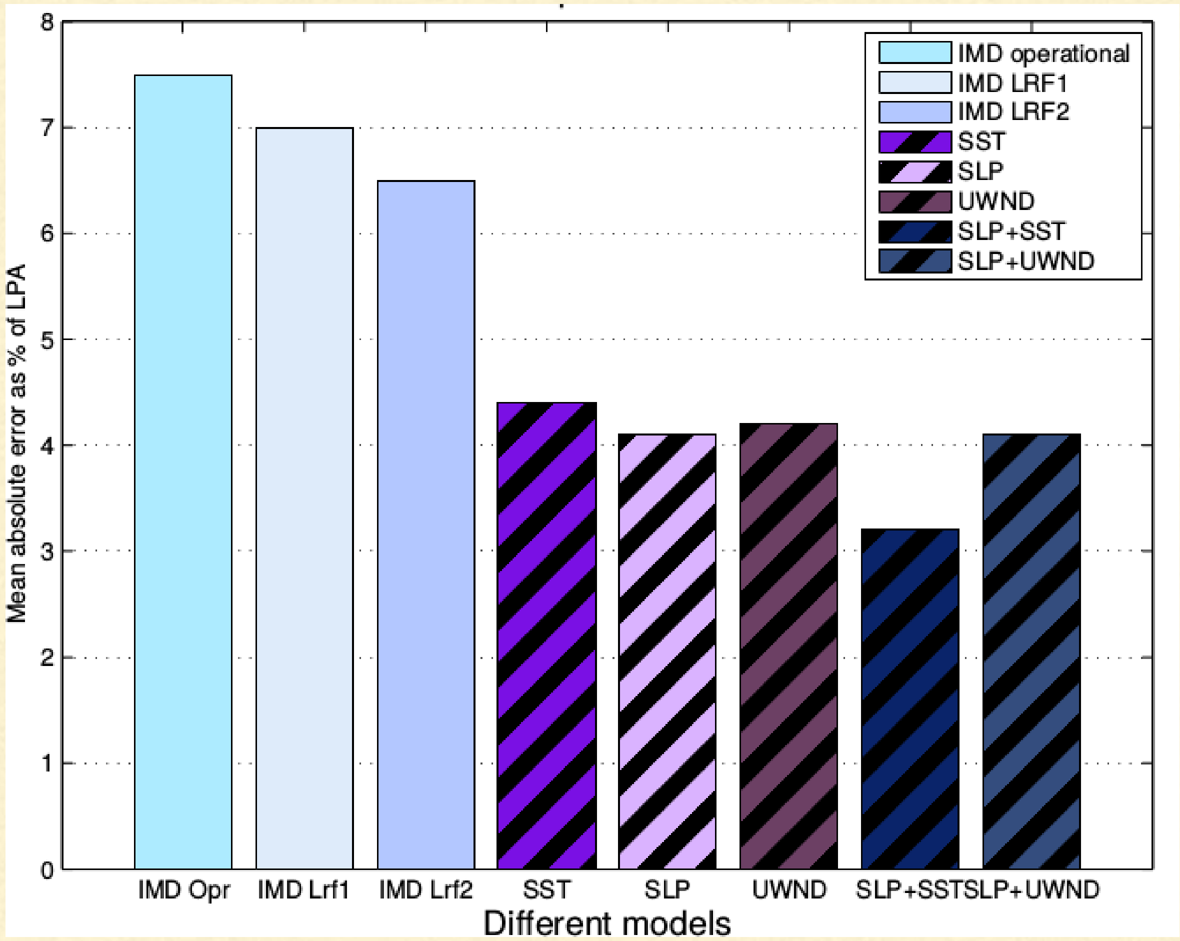
SLP+SST vs individual



SLP+UWND vs individual

Stacked Autoencoder Model vs Existing Model

- IMD operational and PPR models give errors of 7.5%, 7.1%, and 6.5% in May, April, and June
- Combined predictors of SLP+SST and SLP+UWND produce errors of 3.2% and 4.1% in April

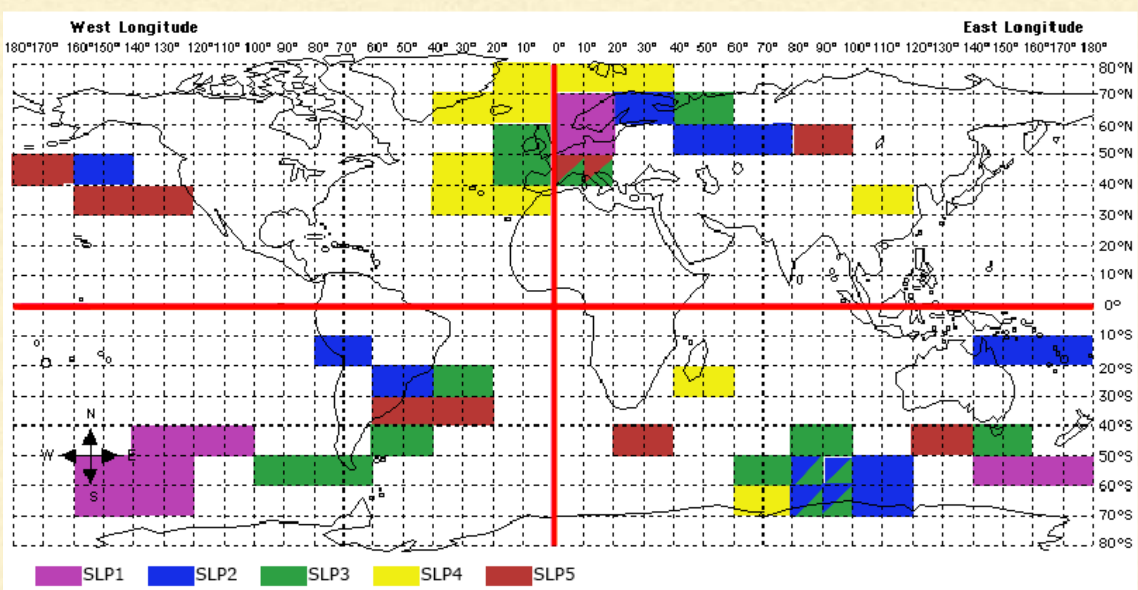


Prediction by stacked autoencoder model and IMD model

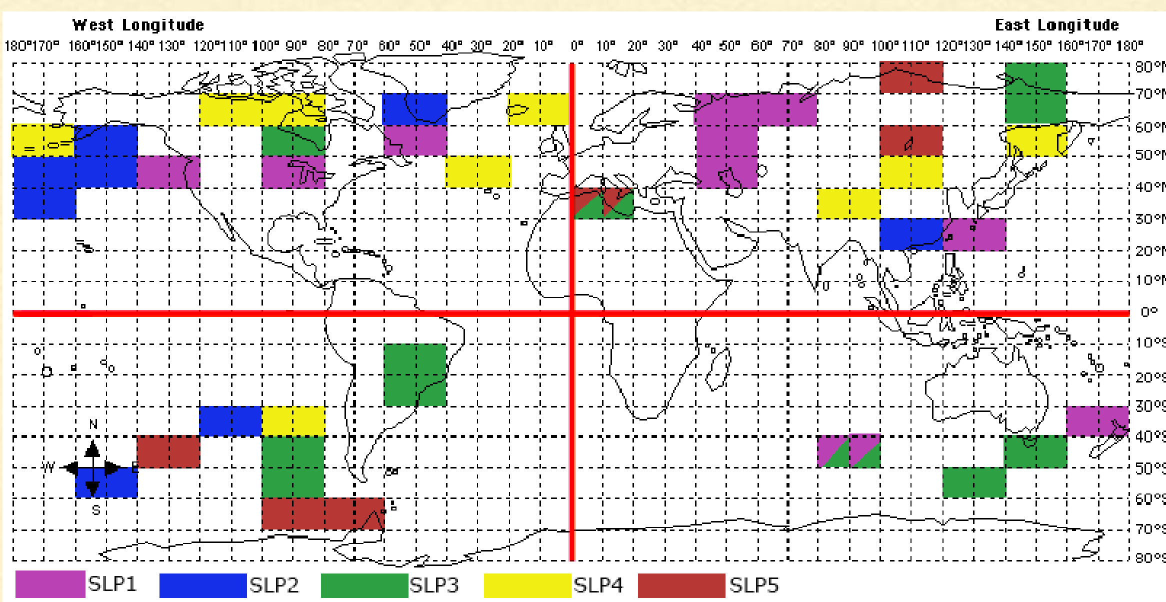
Early and Late Monsoon Predictors

- Spatial coverage of monsoon predictors from SLP for early and late Indian summer monsoons

- Predictors for two phases of rainfalls differ in their locations



Early monsoon predictors



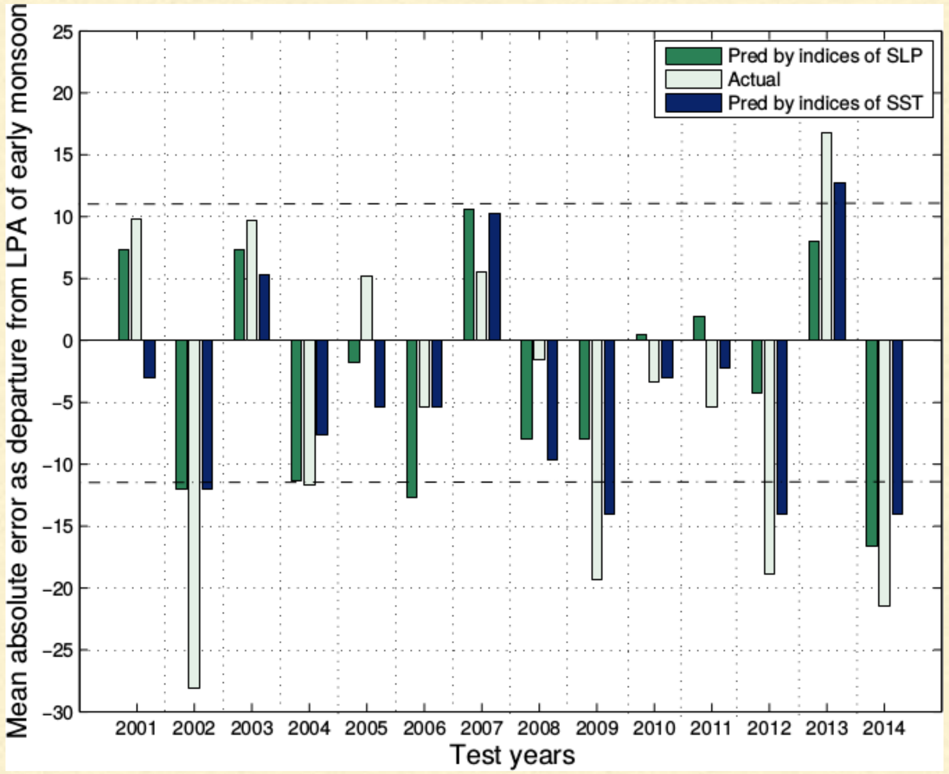
Late monsoon predictors

Prediction Accuracy for Early and Late Monsoon

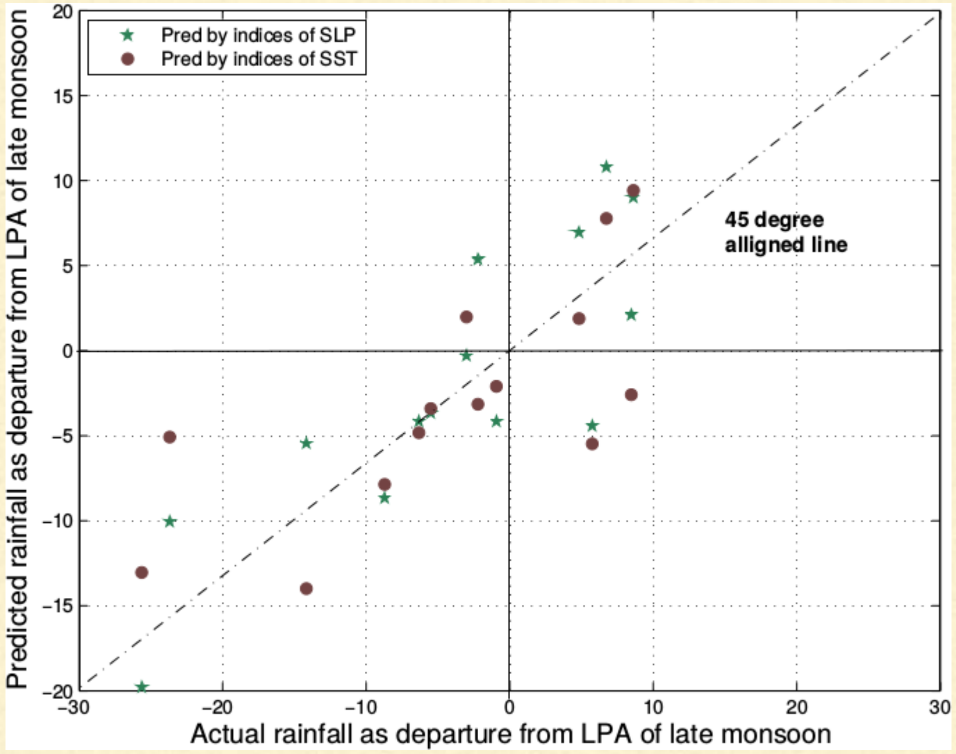
- Forecast early monsoon with 6.1% in April as compared to standard deviation of around 12%

- Error of 4.9% for late monsoon in March in comparison to its standard deviation of 14%

- Prediction of late monsoon is superior to early monsoon



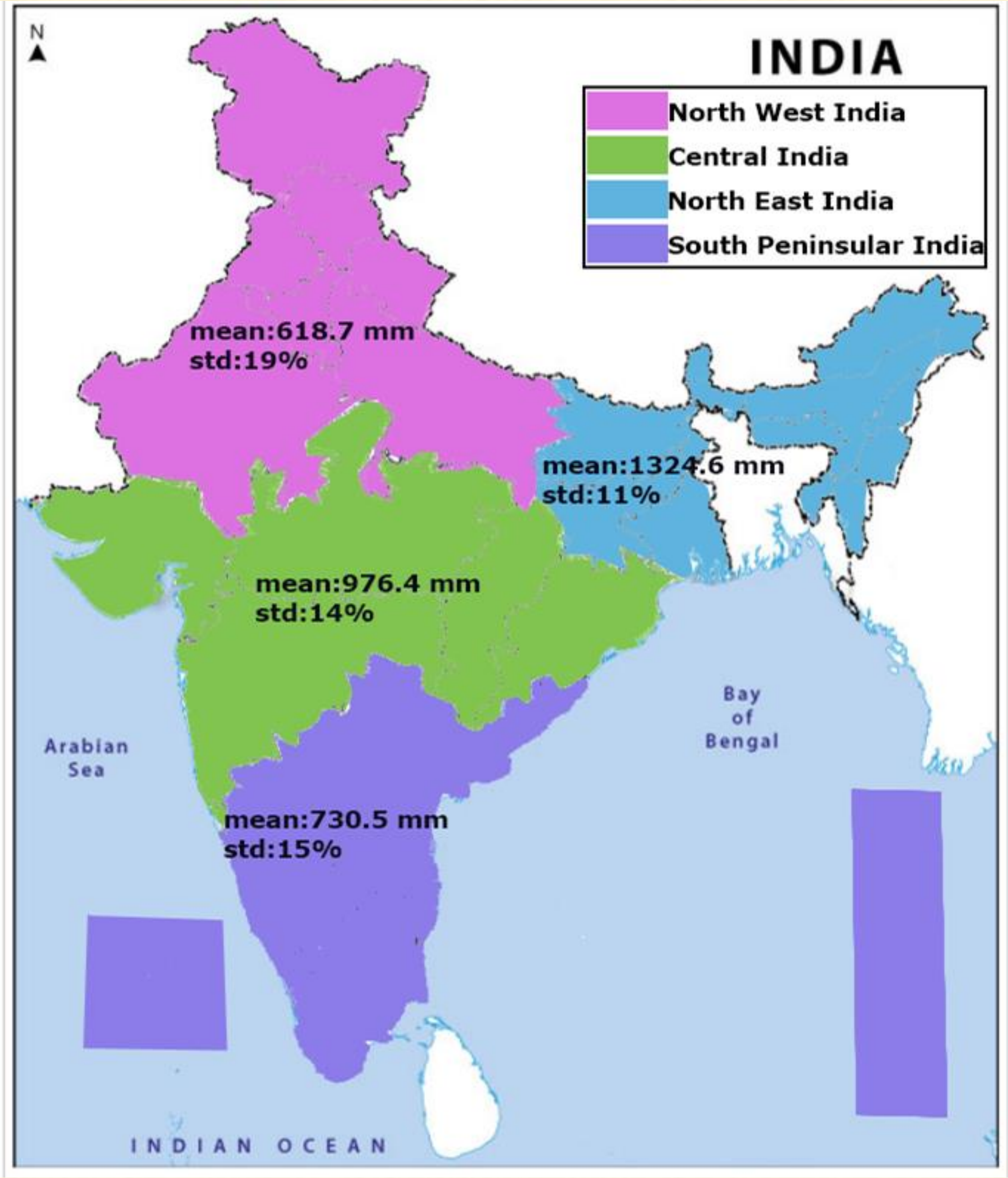
Early monsoon prediction



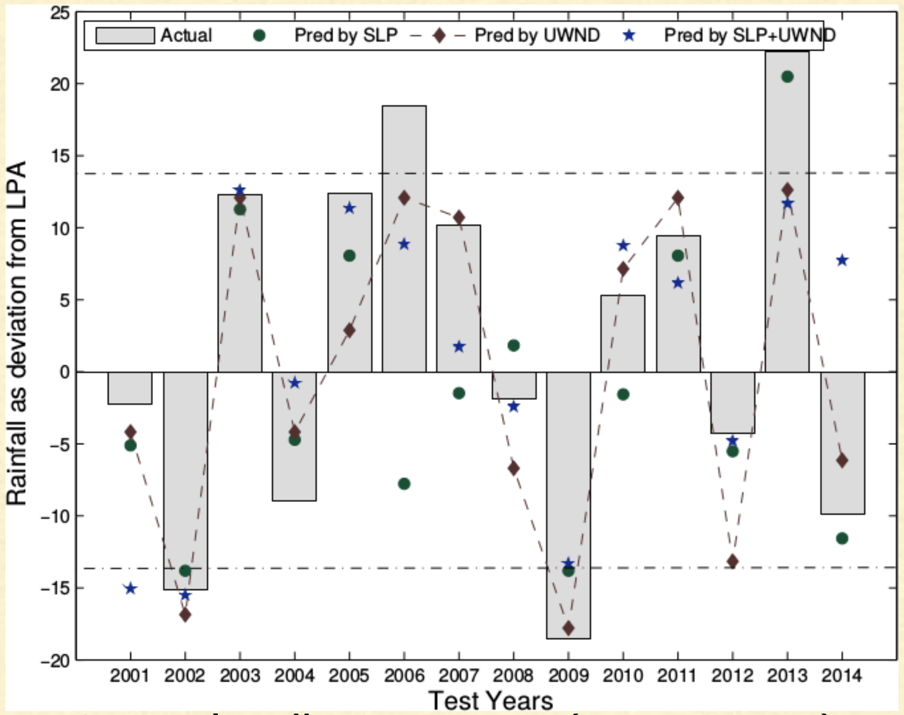
Late monsoon prediction

Homogeneous Regions of India

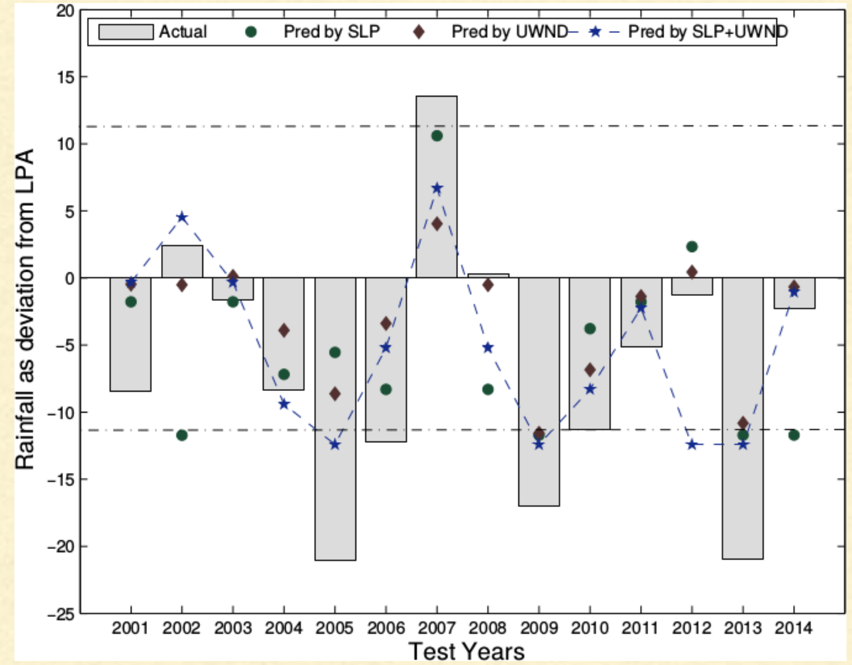
- India Meteorological Department has divided India into four regions following distribution of monsoon



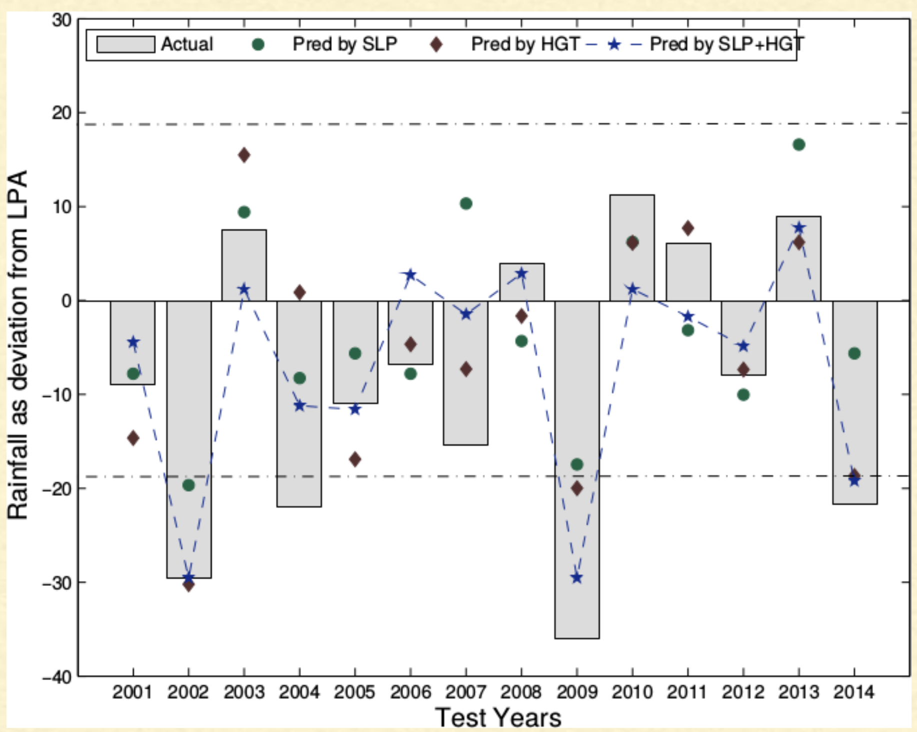
Prediction Accuracy for Regional Indian Monsoon



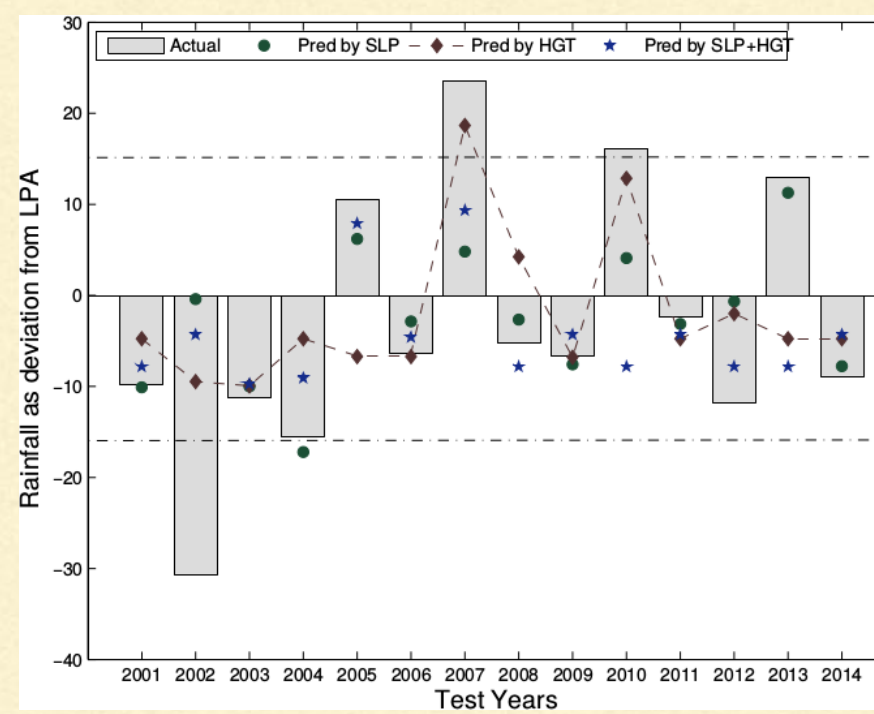
Central India monsoon (Error: 4.1%)



North-east India monsoon (Error: 5.1%)



North-west India monsoon(Error: 5.5%)



South-peninsular India monsoon(Error:6.2%)

Stacked Autoencoder Model vs Existing Model

- Proposed model with identified monsoon predictors outperforms IMD model for all four regions of India

Regional rainfall	IMD model	Month	Proposed model	Month
Central India	12.2	June	**4.8**	January
North-east India	7.8	June	**5.4**	March
North-west India	9.6	June	**6.1**	November
South-peninsular India	8.9	June	**5.3**	March

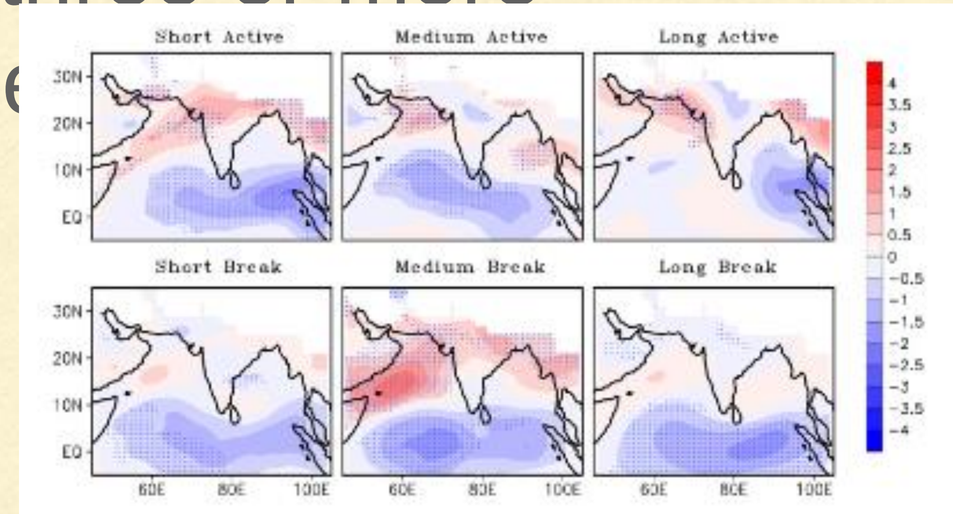
Deep Learning Based LSTM and SeqToSeq Models to Detect Monsoon Spells of India

1. V.Saicharan, M. Saha, P. Mitra, and R. S. Nanjundiah. “Deep Learning Based LSTM and SeqToSeq Models to Detect Monsoon Spells of India ”, ICCS 2018

Active and Break spells

- Significant and challenging to analyze rainfall at daily scale

- Active spells refer to continuous period of three or more days having rainfall above standard de
mean



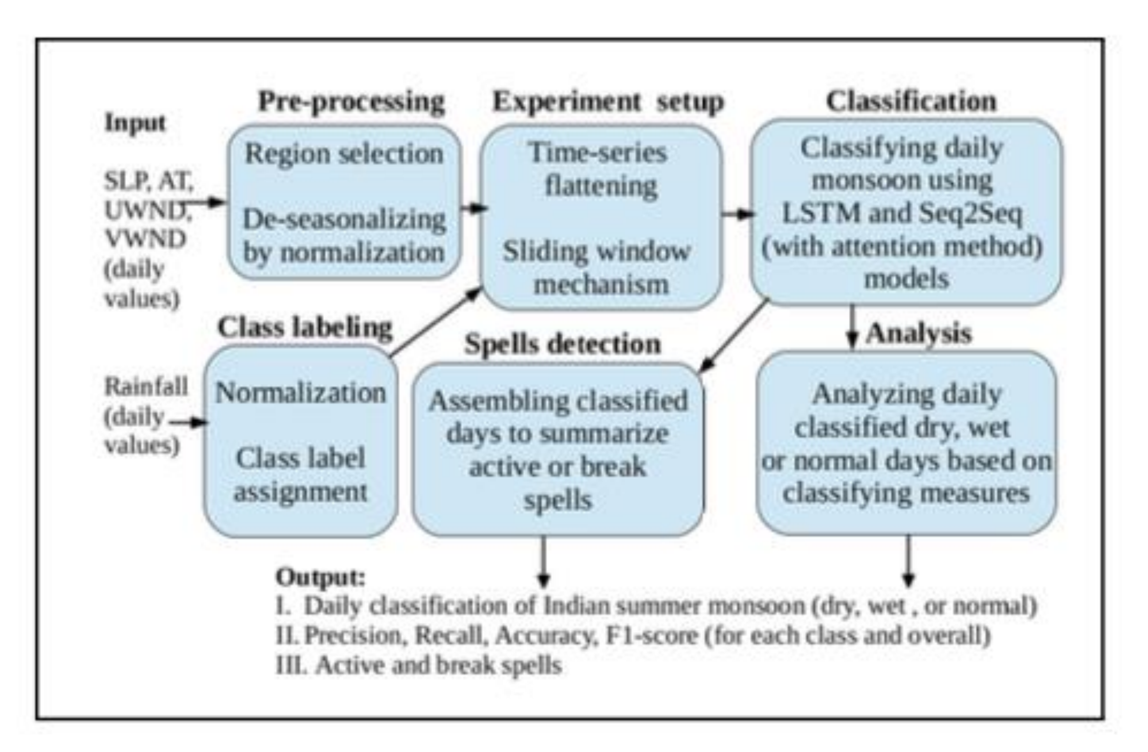
- Break spells refer to the period of three or more days having no rainfall or less than standard deviation from the mean

Active and Break Spells

- Determination of spells assist in proper strategy-building

LSTM Based Prediction of Active and Break spells

- Capable of capturing the long-distance temporal variation and dependencies in data
- Flattening the spatio-temporal input data
- Classifying the days into dry, wet, or normal class usi LSTM or Seq2Seq



Proposed Approach

- Summing up the classified days to detect break or active monsoon spells

Accuracy of Wet and Dry Days Classification

- Prediction results for a lead of one to five days
- Proposed models are compared against conventional Support Vector Machine (SVM), and K-Nearest Neighbour classifiers

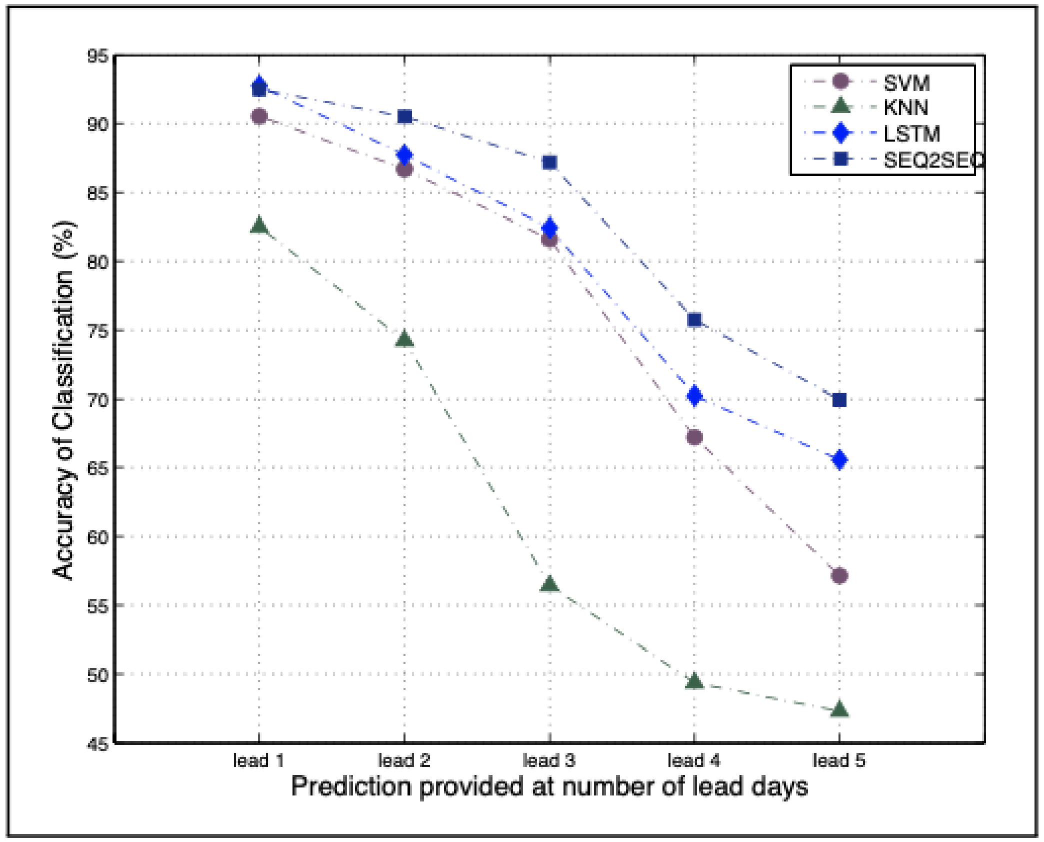
Classification at lead 3

Models	SVM	KNN	LSTM	Seq2Seq
Dry day classification				
Precision	0.793	0.633	0.809	0.909
Recall	0.808	0.596	0.840	0.834
Wet day classification				
Precision	0.784	0.319	0.782	0.796
Recall	0.672	0.252	0.652	0.744
Normal day classification				
Precision	0.871	0.741	0.882	0.911
Recall	0.897	0.781	0.901	0.932
Overall classification				
Accuracy	81.60	56.43	82.43	87.20
F1-score	0.849	0.667	0.859	0.890

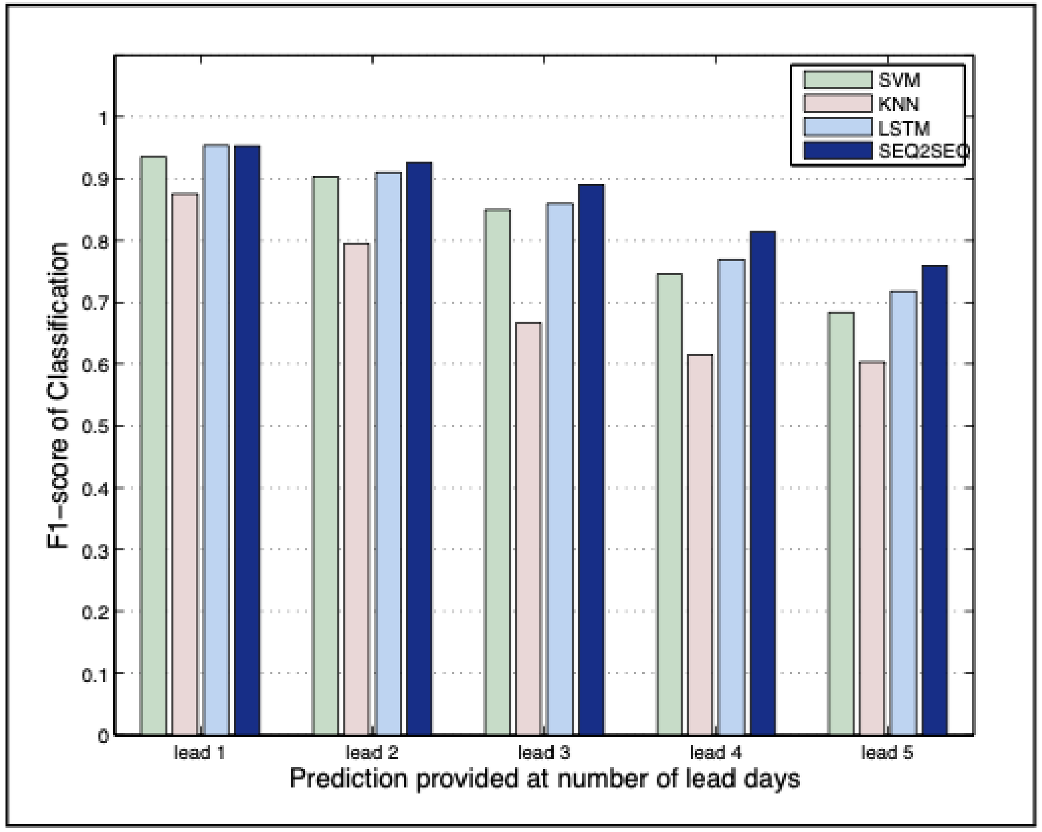
LSTM and Seq2Seq models vs Traditional Classifiers

- Proposed Seq2Seq and LSTM models outperforms conventional SVM and KNN classifiers

- Accuracy decreases as lead increases from one to five days



Accuracy of classification

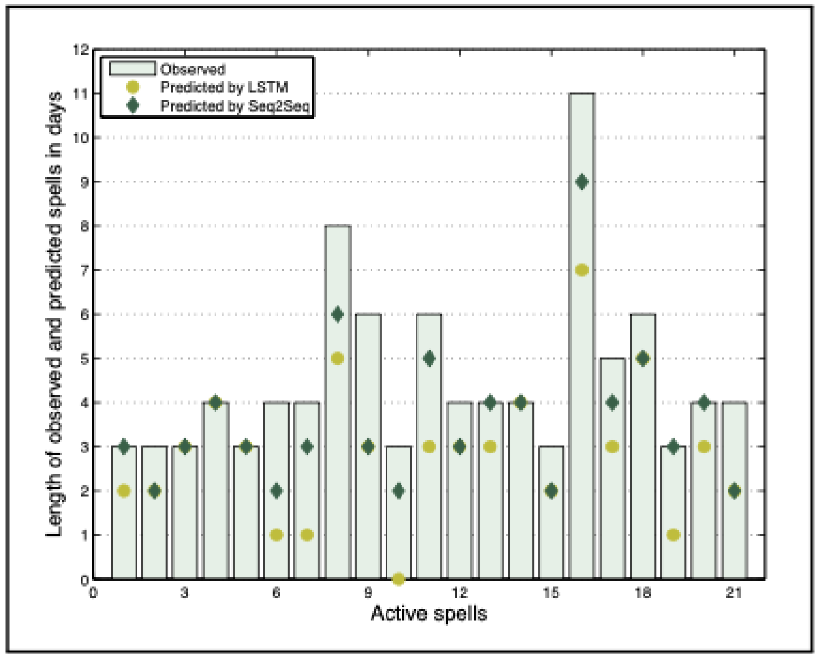


F1-score of classification

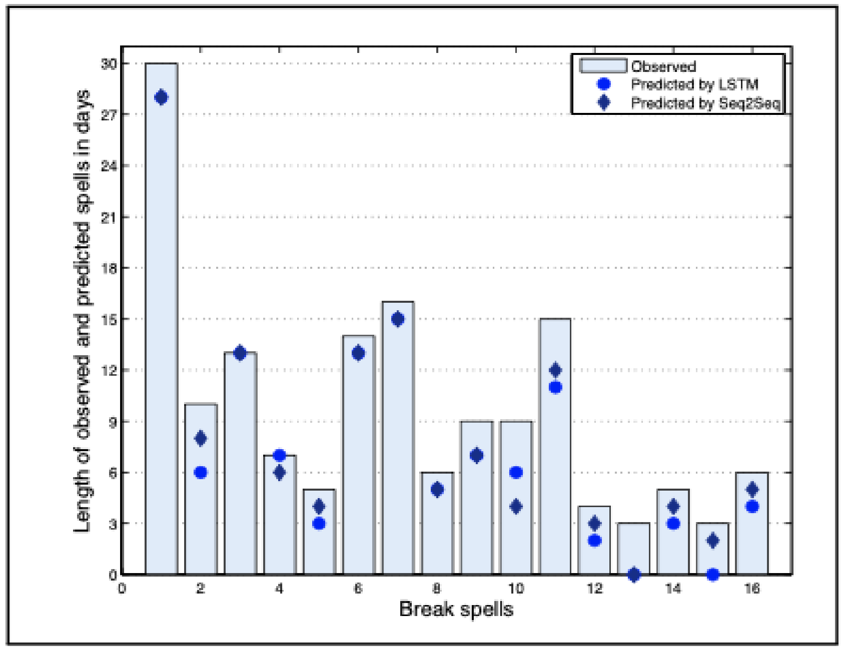
Prediction of Monsoon Spells

- Classified dry and wet days are summed to identify the spells
- Seq2Seq predicted 14/16 break and 16/21 active spells

Models	Observed # of break spells	Predicted # of break spells	Observed # of active spells	Predicted # of active spells
LSTM	16	13	21	13
Seq2Seq	16	14	21	16



Obs vs Pred active spells



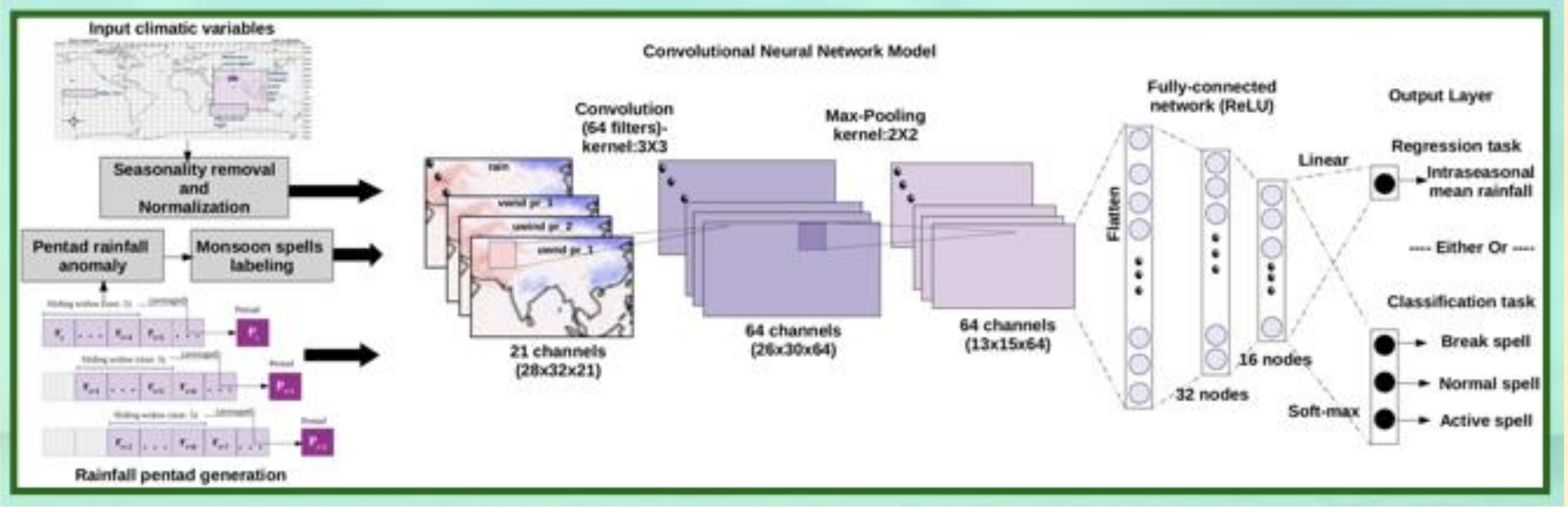
Obs vs Pred break spells

CNN-Based Forecasting of Intraseasonal Mean and Active/Break Spells for Indian Summer Monsoon

1. M. Saha, R.S. Nanjundiah and C. Monteleoni. “CNN-Based Forecasting of Intraseasonal Mean and Active/Break Spells for Indian Summer Monsoon”, CI 2020
2. M. Saha, R. S. Nanjundiah, and C. Monteleoni. “Prediction of Intraseasonal Mean and Active/Break Spells for Indian Summer Monsoon Using a CNN”, WiML-ICML 2020
3. M. Saha, B. Finley and C. Monteleoni. “Deep Convolutional Network for Classifying Indian Summer Monsoons at a Daily Scale”, WiML-NeurIPS 2019

Convolutional Neural Network Based Prediction of Active/Break Spells

- CNN assists considering the spatial relationships between the climatic variables
- Earlier time-steps and different pressure-levels are added as channels of the input and that considers the time-dependency



Classification of Daily Indian Monsoon

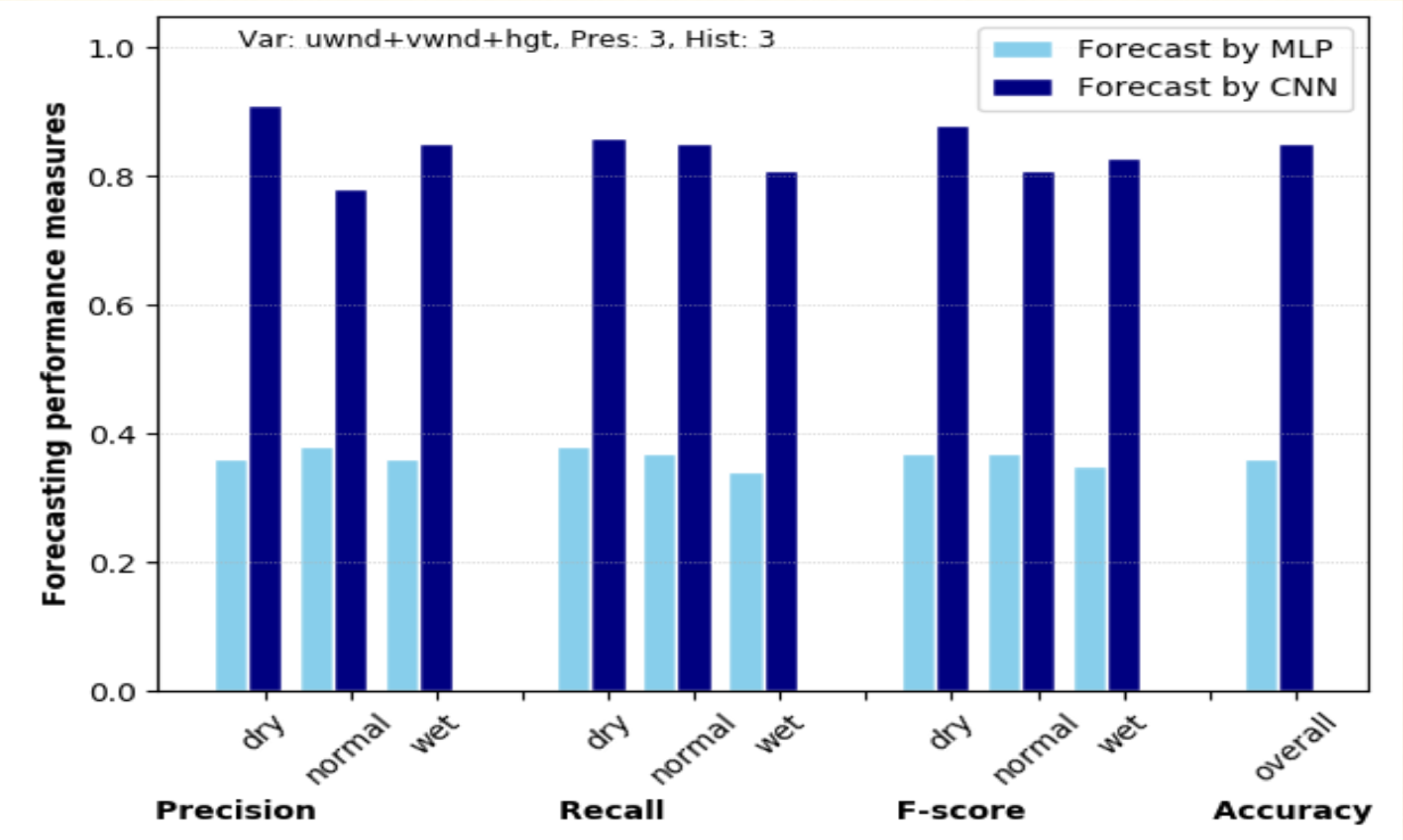
- Forecasting dry and wet days is a vital as they may results in drought or flood
- CNN trained with a combination of three variables shows higher accuracy than model with individual

Tab 1: Forecast of daily Indian rainfall using CNN model

	Precision	Recall	F1-Score
Dry day	0.91	0.86	0.88
Normal day	0.78	0.85	0.81
Wet day	0.85	0.81	0.83
Overall accuracy	0.85		

CNN Model vs Multilayer Perceptron

- CNN shows significant improvement over MLP model

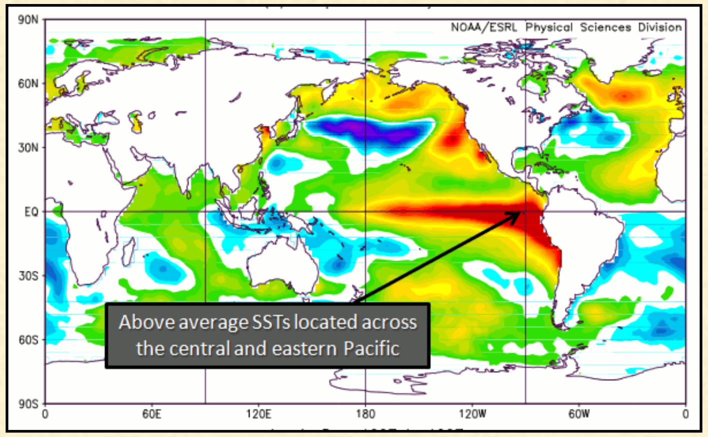


Deep Learning Based Prediction of ENSO and EQUINOO Indices during June-September

1. M. Saha and R. S. Nanjundiah. “Prediction of ENSO and EQUINOO In- dices during June to September using Deep Learning Method”, Meteorological Applications, 2020

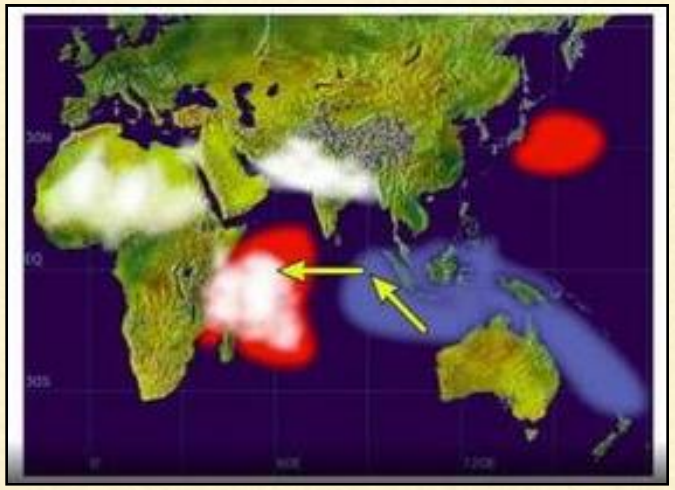
ENSO and EQUINOO

- ENSO: resembles the irregularly repeated alteration in atmospheric winds and sea surface temperature over tropical eastern Pacific Ocean



ENSO

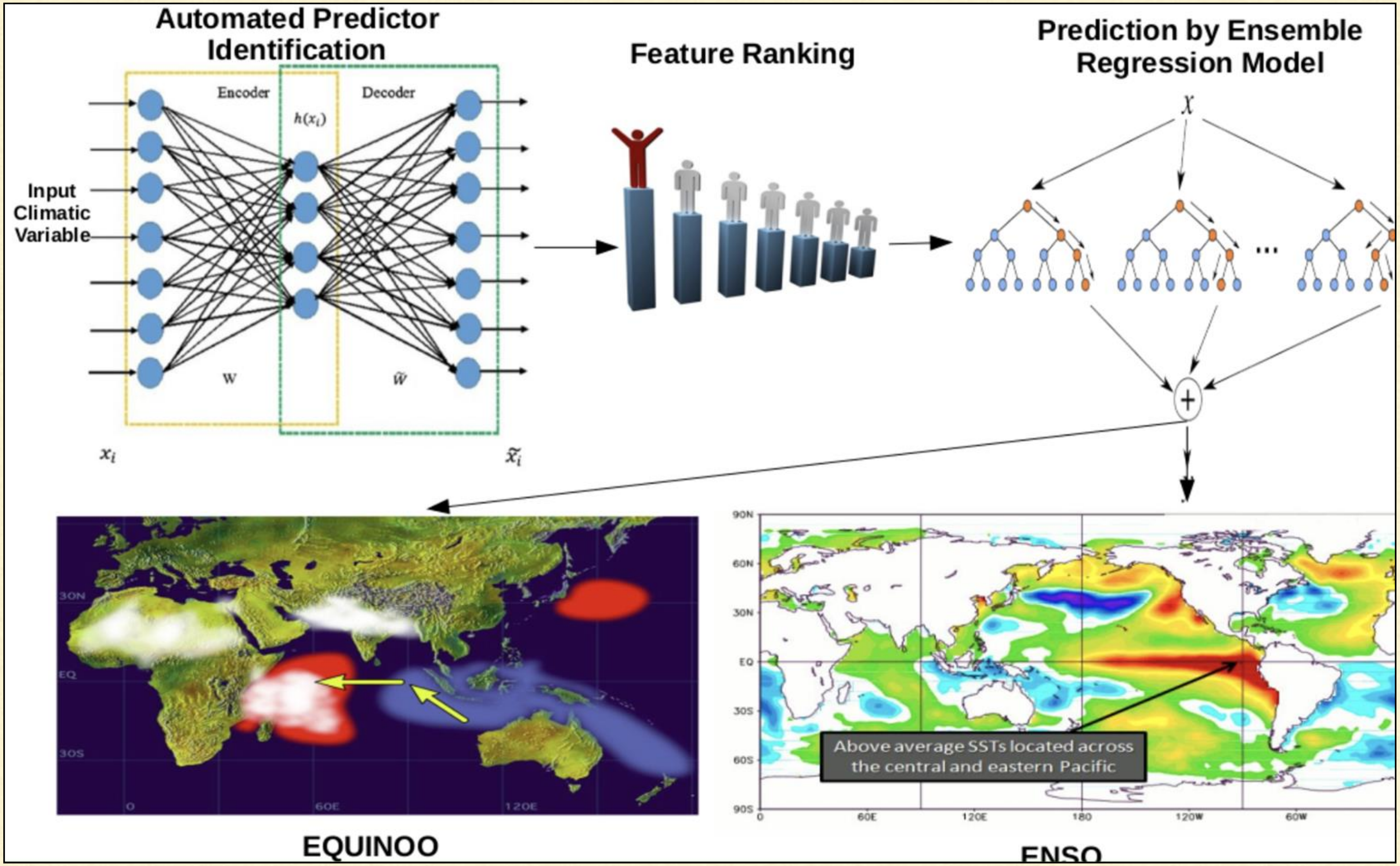
- EQUINOO: Oscillation in convection between Western and Eastern equatorial Indian Ocean



EQUINOO

- ENSO and EQUINOO indices influence multiple climatic phenomenon including Indian summer monsoon

Autoencoder-Based Prediction of ENSO and EQUINOO Indices



Prediction of ENSO

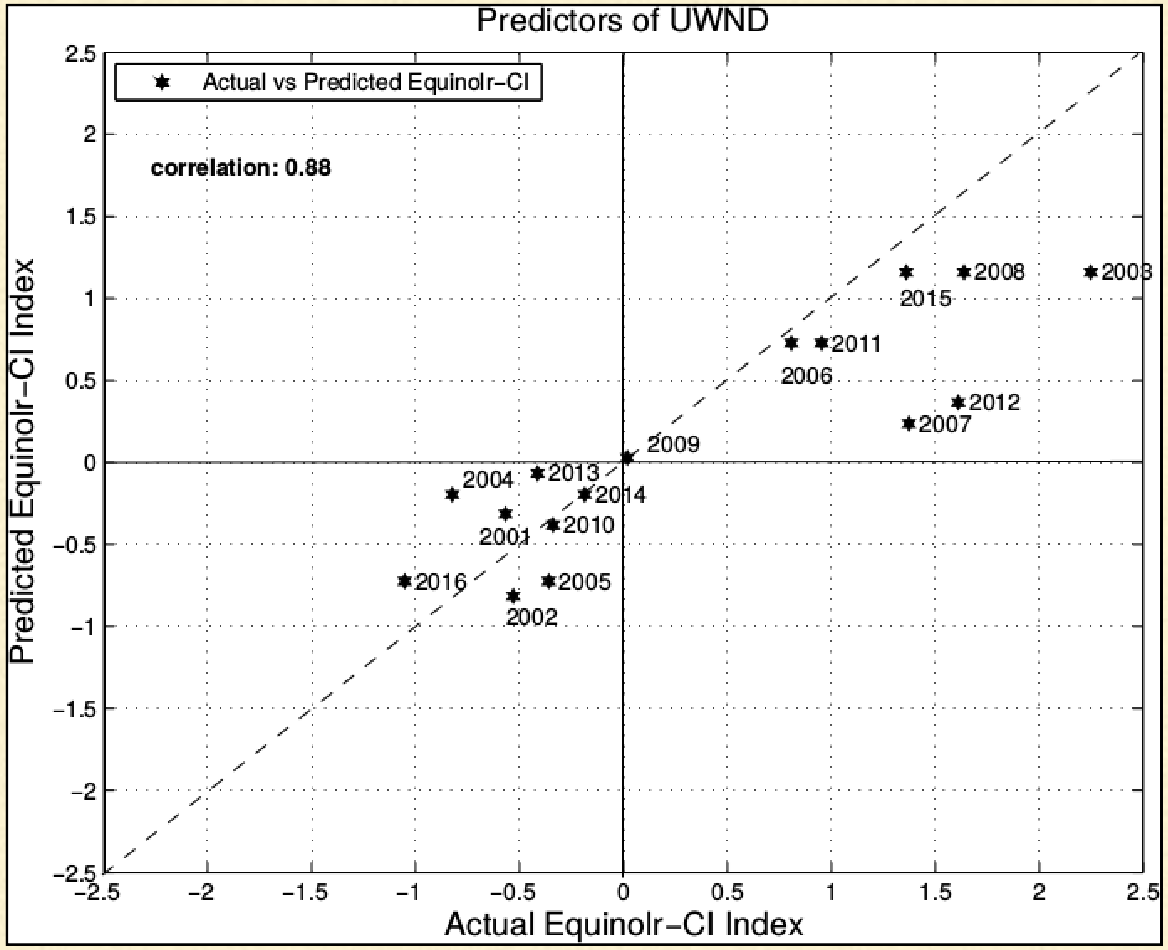
- Predicted ENSO for JJAS with correlation of 0.87

Measures for ENSO	Values				
	JJAS	June	July	Aug.	Sep.
Correlation	0.87	0.88	0.88	0.84	0.87
Sensitivity	0.77	1.0	0.88	0.75	0.87
Specificity	0.85	1.0	1.0	0.87	0.87
Precision	0.87	1.0	1.0	0.85	0.87
Neg. pred. rate	0.75	1.0	0.87	0.77	0.87
Accuracy	81.2	100	93.7	81.2	87.5
F1 score	0.82	1.0	0.94	0.80	0.87

Prediction of EQUINOO

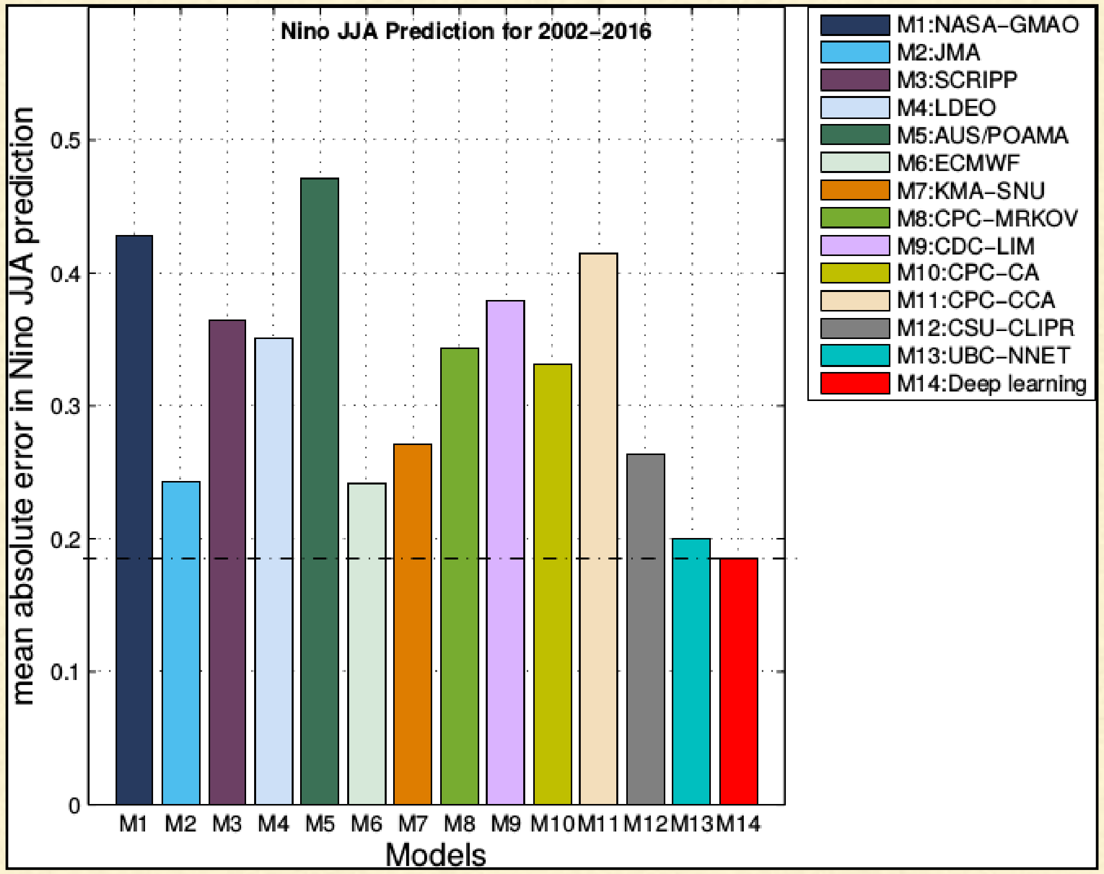
- Predicted EQUINOO with a Pearson correlation of 0.94 for June

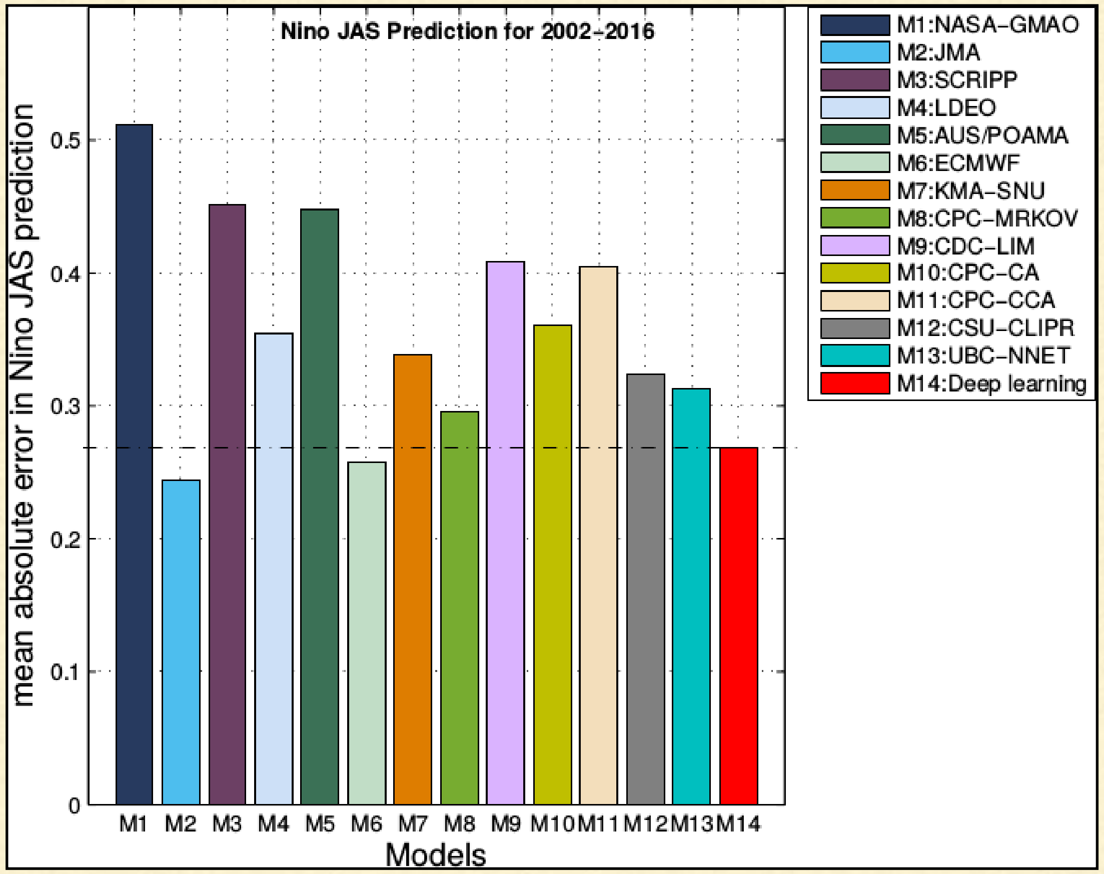
Measures for EQUINOO	Values				
	JJAS	June	July	Aug.	Sept.
Correlation	0.88	0.94	0.86	0.91	0.84
Sensitivity	1.0	0.85	0.80	1.0	0.90
Specificity	1.0	1.0	1.0	0.87	0.83
Precision	1.0	1.0	1.0	0.88	0.90
Neg. predictive rate	1.0	0.90	0.75	1.0	0.83
Accuracy	100	93.7	87.5	93.7	87.5
F1 score	1.0	0.92	0.88	0.94	0.90

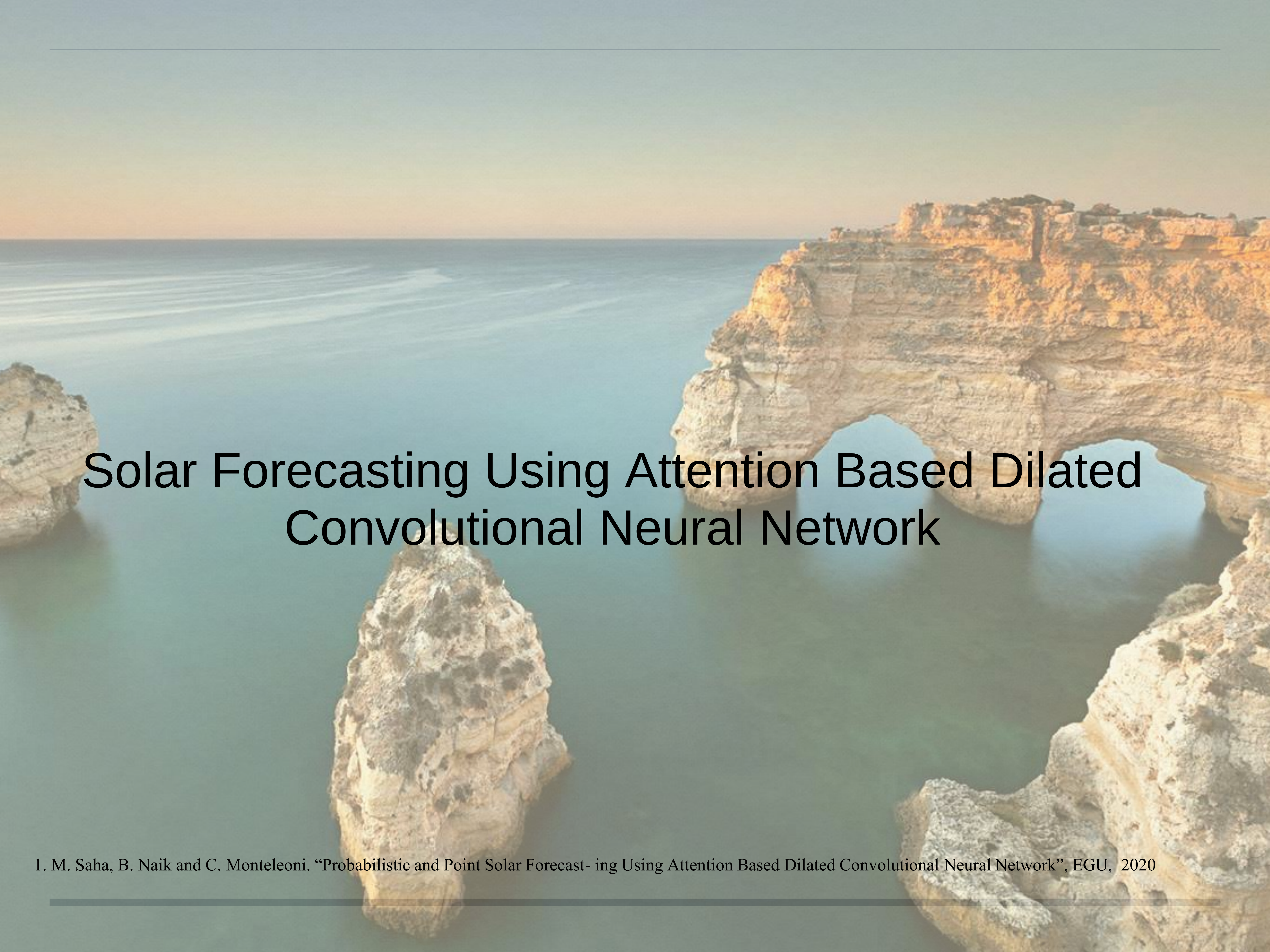


Autoencoder Model vs Existing Model

- Compared with thirteen existing ENSO prediction model
- Proposed approach performs best for June-July-August NINO





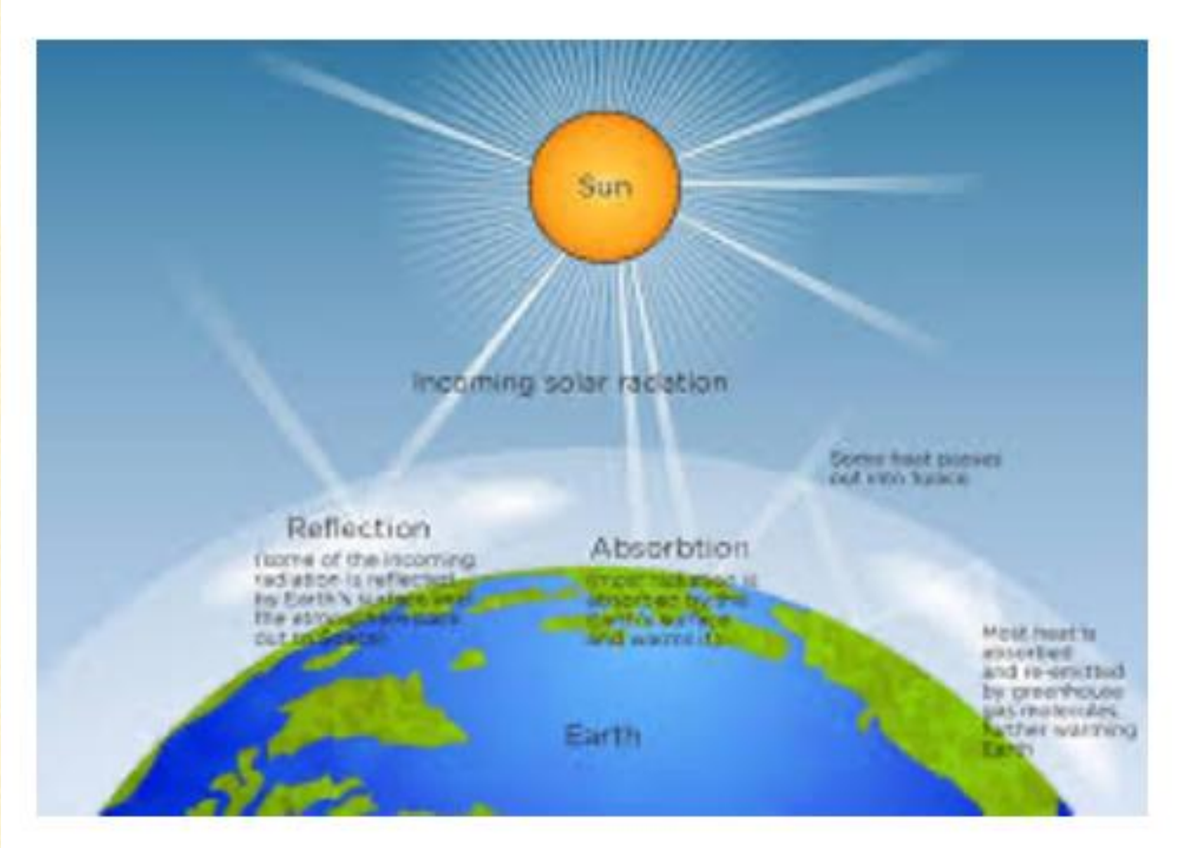


Solar Forecasting Using Attention Based Dilated Convolutional Neural Network

1. M. Saha, B. Naik and C. Monteleoni. “Probabilistic and Point Solar Forecast- ing Using Attention Based Dilated Convolutional Neural Network”, EGU, 2020

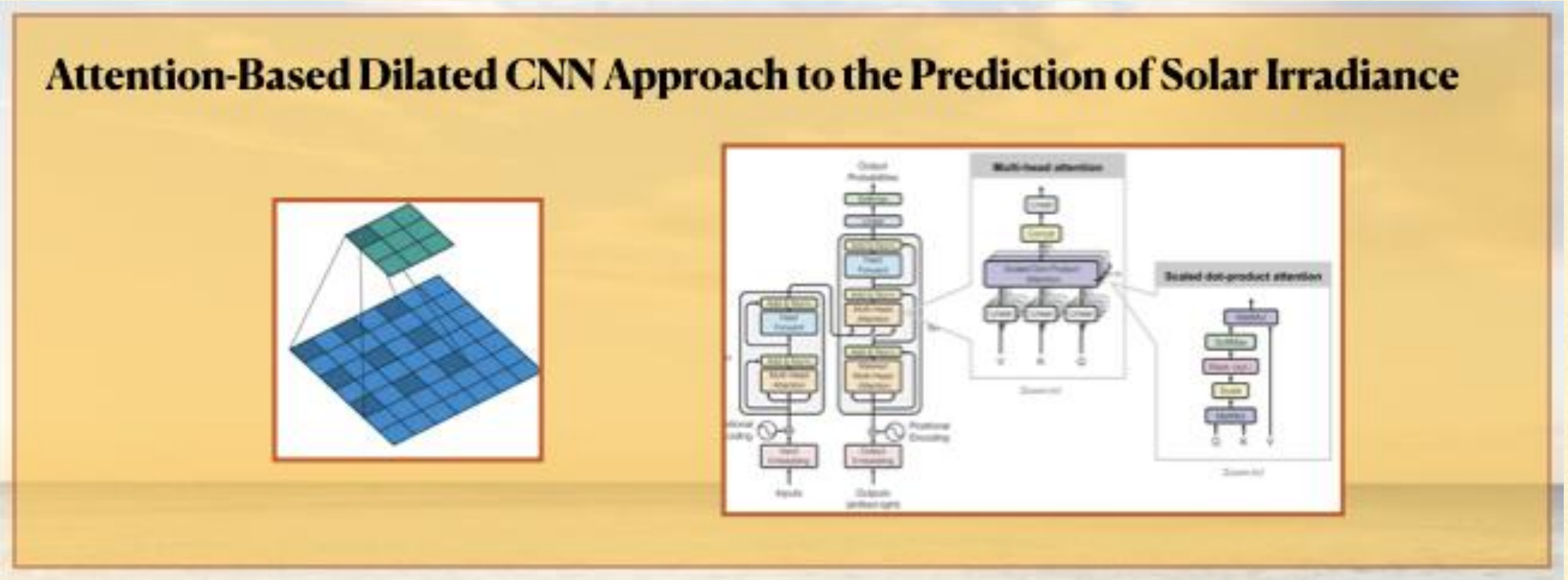
Solar Irradiance

- Solar is a good source for renewable and clean energy
- Solar Irradiance is the flux of radiant energy received per unit area of the earth
- Solar irradiance has many significant applications:
    - the prediction of energy generation
      from solar power plant
    - the heating and cooling loads of buildings
    - climate modelling and weather forecasting



CNN-Based Method with Added Attention

- Convolutional neural networks (CNN) are capable of extracting features from data that have local spatial relations
- We added dilation to the CNN kernel for capturing long-term dependencies
- Attention mechanism compels the model to focus on the parts of the input that bear a high impact on the output



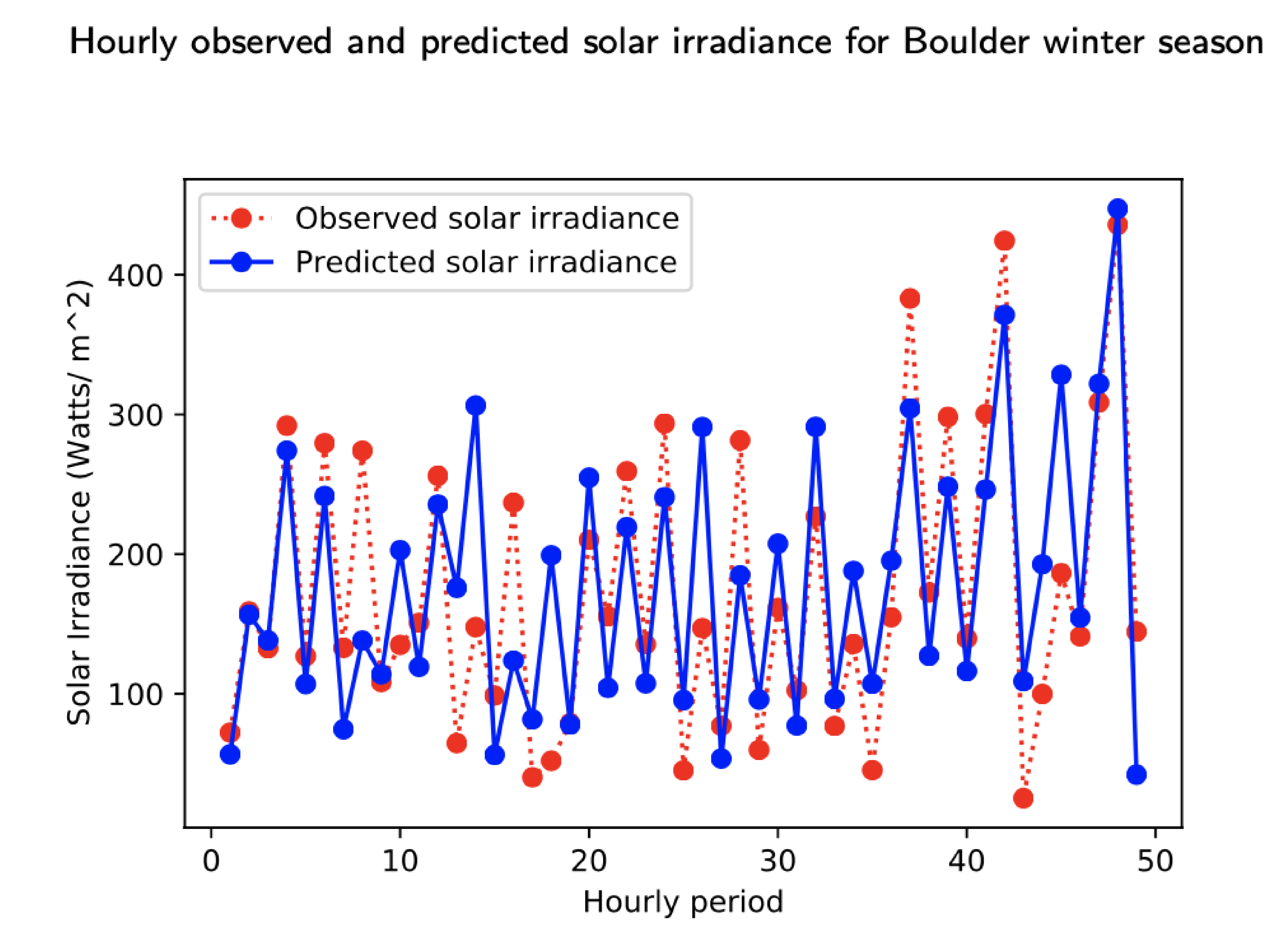
Forecasting of Solar Irradiance

- Forecast is provided for all four seasons at two different leads of 3 and 6 hours
- The model shows higher performance for the fall and winter seasons

RMSE for point solar irradiance forecasting by CNN and simple persistence (SP) models at two different leads

Boulder-Colorado								
Leads	**Fall**		**Winter**		**Spring**		**Summer**	
	CNN	**SP**	**CNN**	**SP**	**CNN**	**SP**	**CNN**	**SP**
3 hrs	169	325	122	280	234	369	243	375
6 hrs	183	375	100	263	267	464	238	491
Fort Peck-Montana								
3 hrs	135	248	128	201	167	304	202	326
6 hrs	148	279	132	174	195	392	252	383

Observed vs Predicted Solar Irradiance



CNN Model vs Persistence Model