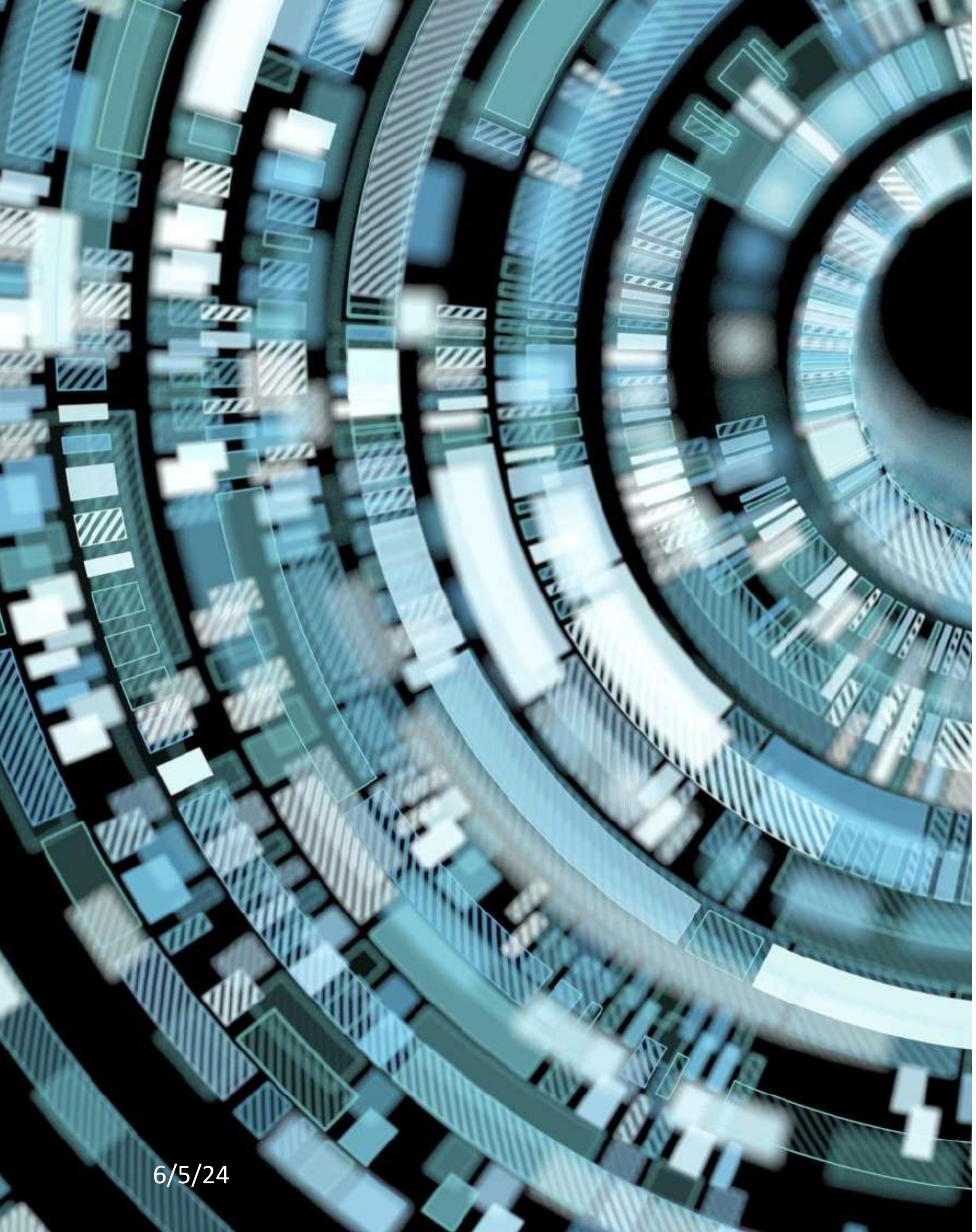
Deep-learning based downscaling approaches for precipitation extremes: an assessment over India

Pankaj Kumar

Associate Professor

Indian Institute of Science Education and Research (**IISER**) Bhopal



Acknowledgement to Group Members

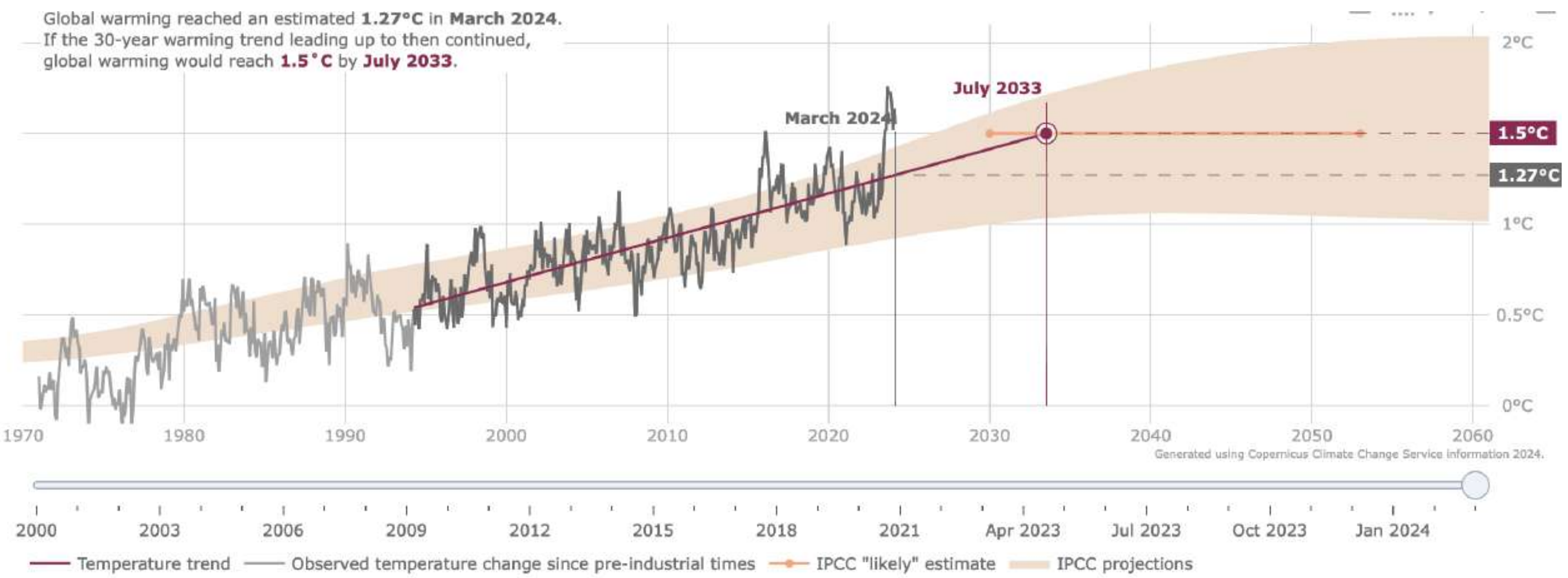
Midhun M & **Sreevastha Golla**

Downscaling and reconstruction of high-resolution gridded rainfall data over India using deep learning-based generative adversarial network. ***Modeling Earth Systems and Environment (2023)***

https://doi.org/10.1007/s40808-023-01899-9

How close are we to reaching a global warming of 1.5°C?

- According to IPCC AR6, the temperature over the land has risen to 1.59 °C [1.34 to 1.83] compared to the 1850-1900 average w.r.t. 2011-2020
- Human-induced GHG forcing is the main driver of the observed changes in hot and cold extremes on the global scale (virtually certain) and on most continents (very likely).
- Climate models are capable of addressing the science of climate change but do show limitations at fine scales to capture extremes.
- The emerging AI/ML techniques have shown a potential to address it on a sub-grid scale.



Earth's global surface temperature relative to the pre-industrial period will be 1.5^{0}C in July 2033

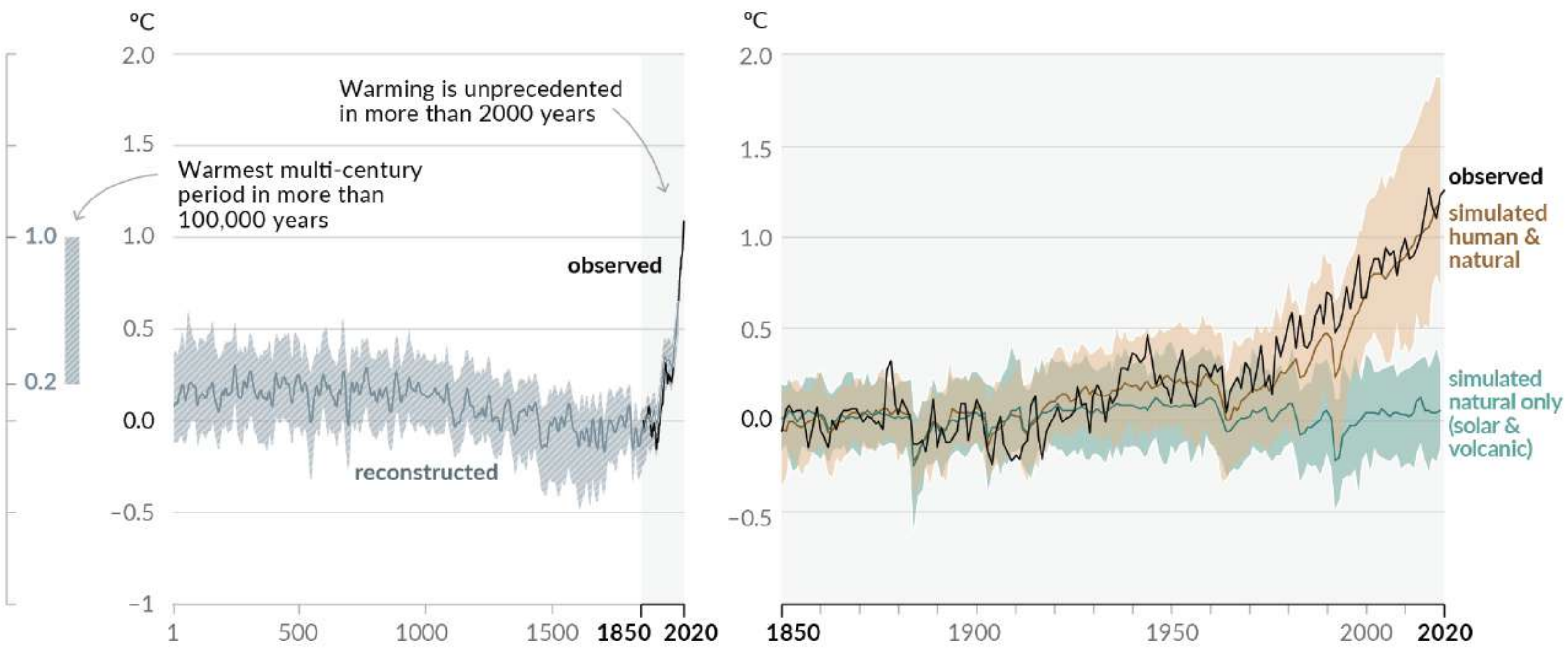
Unprecedented Warming and Increasing Extremes

Human influence has warmed the climate at a rate that is unprecedented in at least the last 2000 years

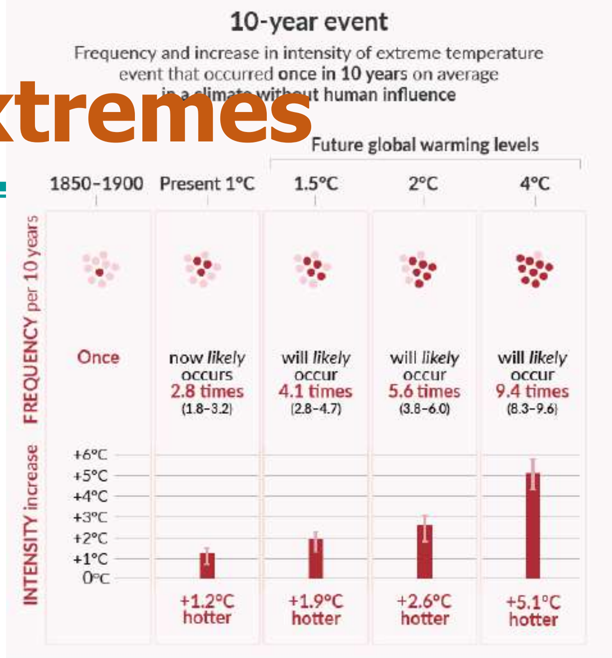
Changes in global surface temperature relative to 1850–1900

(a) Change in global surface temperature (decadal average) as reconstructed (1–2000) and **observed** (1850–2020)

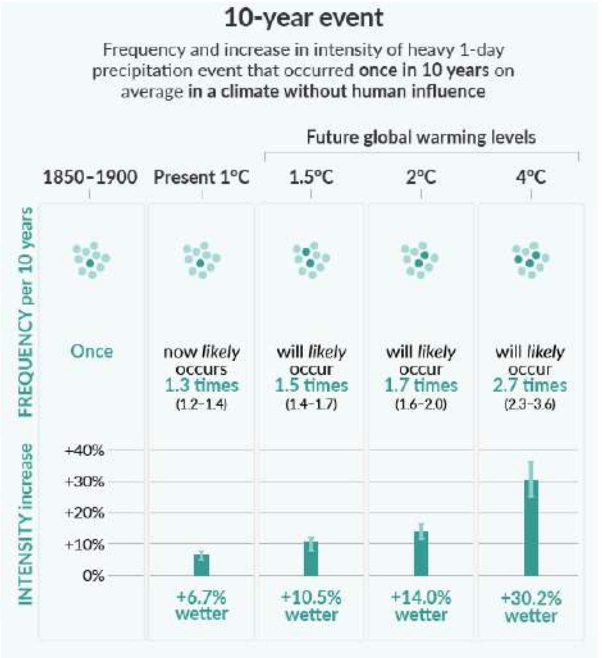
(b) Change in global surface temperature (annual average) as **observed** and simulated using human & natural and only natural factors (both 1850–2020)



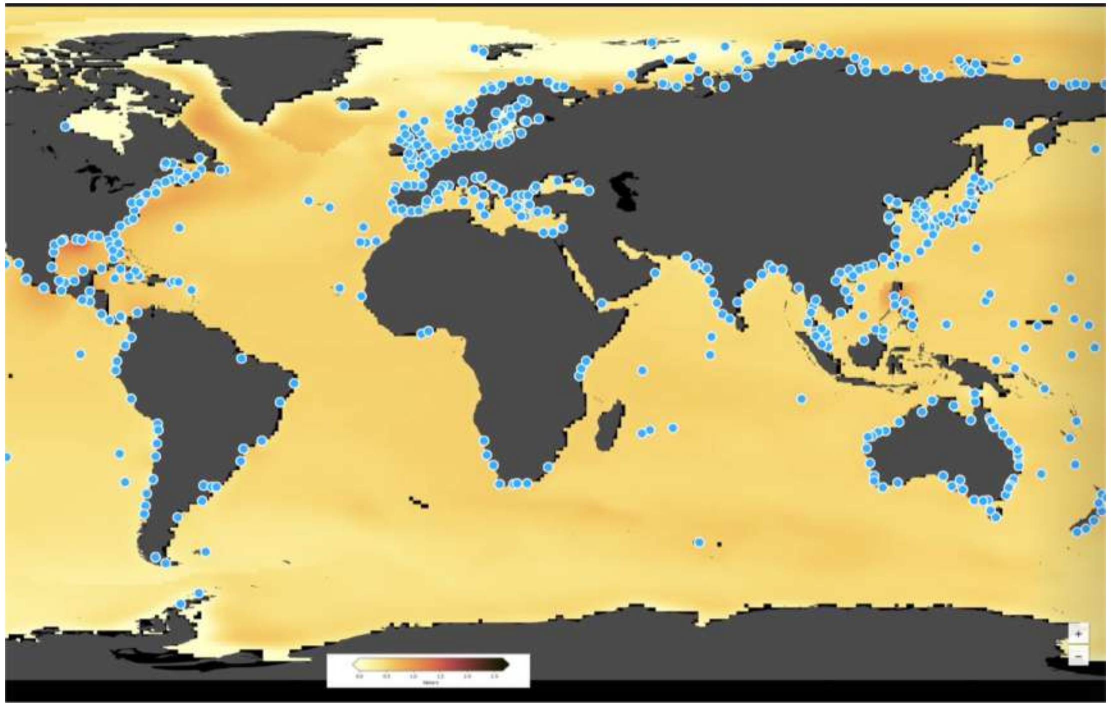
Heatwave



Heavy Precipitation



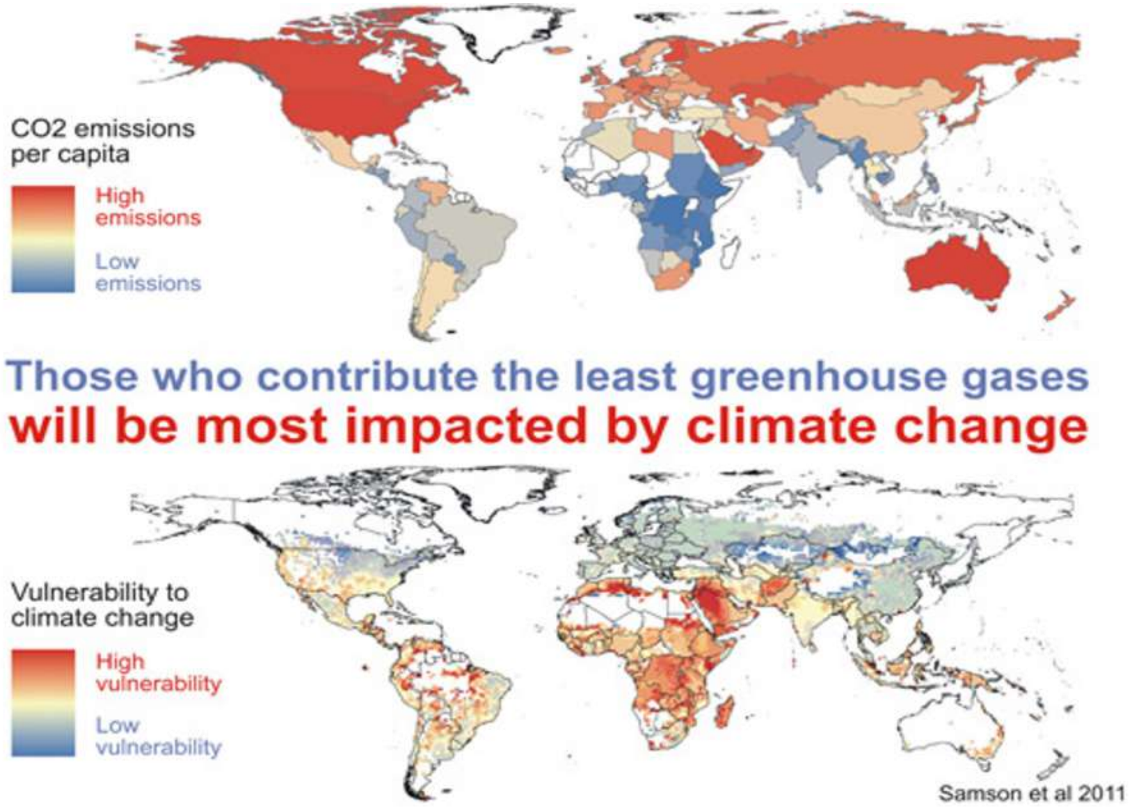
Future Projections



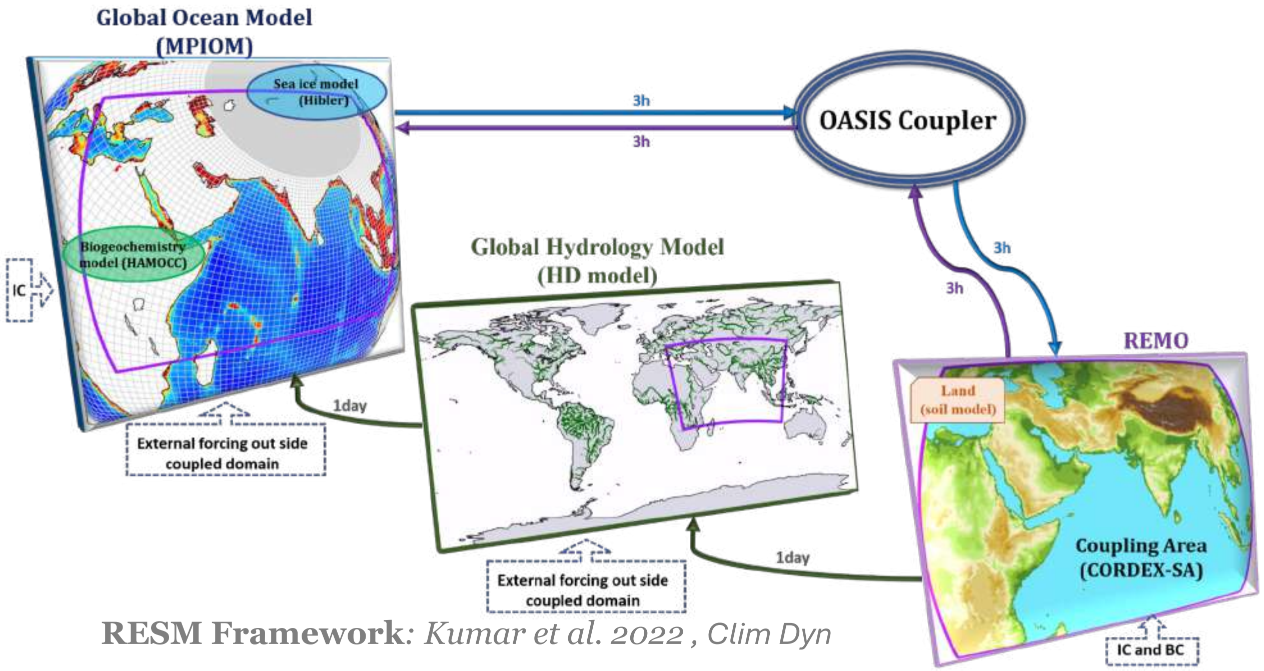
The project accelerated warming will lead to significant **sea-level rise**, and it is projected by the end of this century at least **12 coastal Indian cities water levels will increase by 2-3 feet**, due to climate change

Source: IPCC AR6

Global Vulnerability of Climate Change

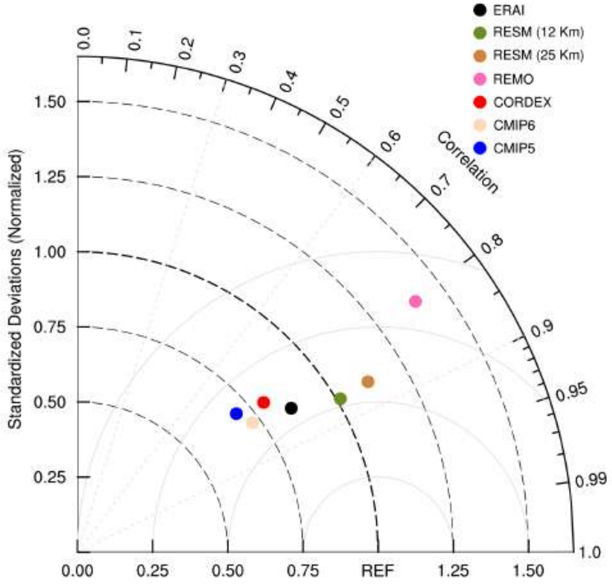


@ Regional Earth System Modelling Framework (RESM): CORDEX-SA



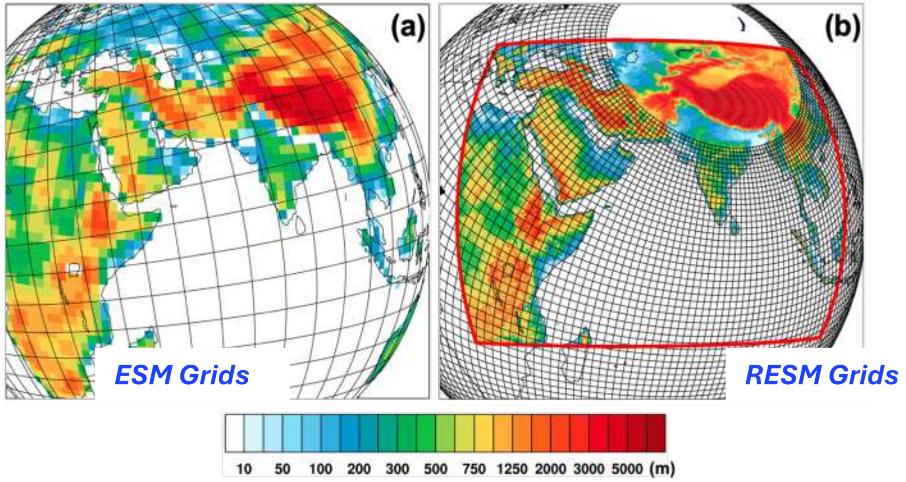
RESM Framework: *Kumar et al. 2022 , Clim Dyn*

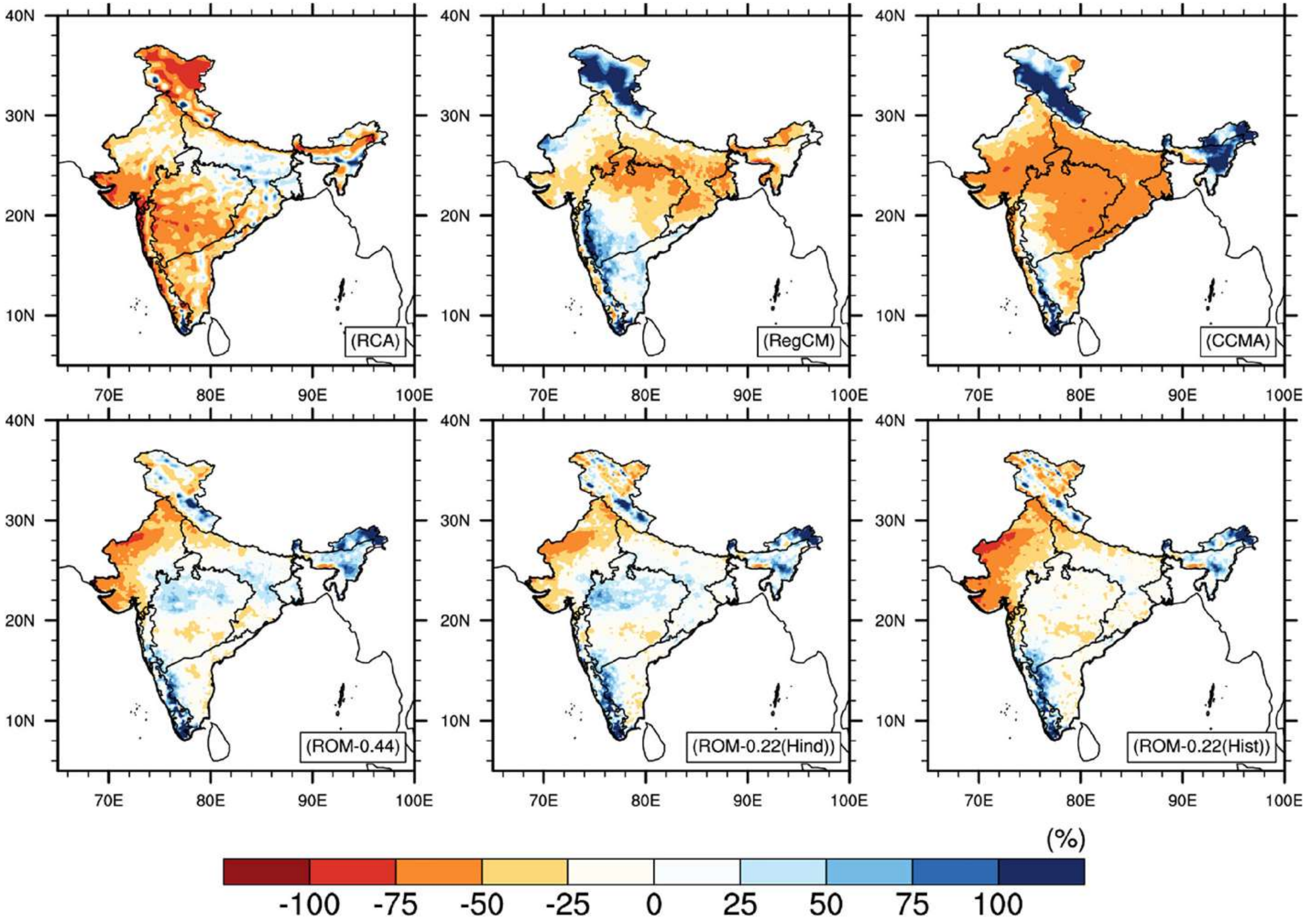
RESM Components	Models	Experiments	
Atmospheric	REMO (2009) (regional)	Horizontal Resolution	**0.22⁰/0.11⁰**
		Vertical levels	27
		Boundary conditions	ERA-I, CMIP5/6
Ocean	MPIOM (global)	Horizontal Resolution	**~5 to ~25 km** over **IO**
		Vertical levels	40
Marine Biogeochemistry	HAMOCC (global)		
Hydrological Discharge	HD (global)	Reanalysis	0.5^0
Coupling frequency	MPIOM & REMO: 3 hourly; HD model: 1 day		
Simulation period	**Hist**: 1960-2014; **Scenarios**:1960-2100, excluding spin-up		



RESM showed **high skills** compared to other suites of models over the **CORDEX-SA** region.

Kumar et al. 2022, Atmospheric Research

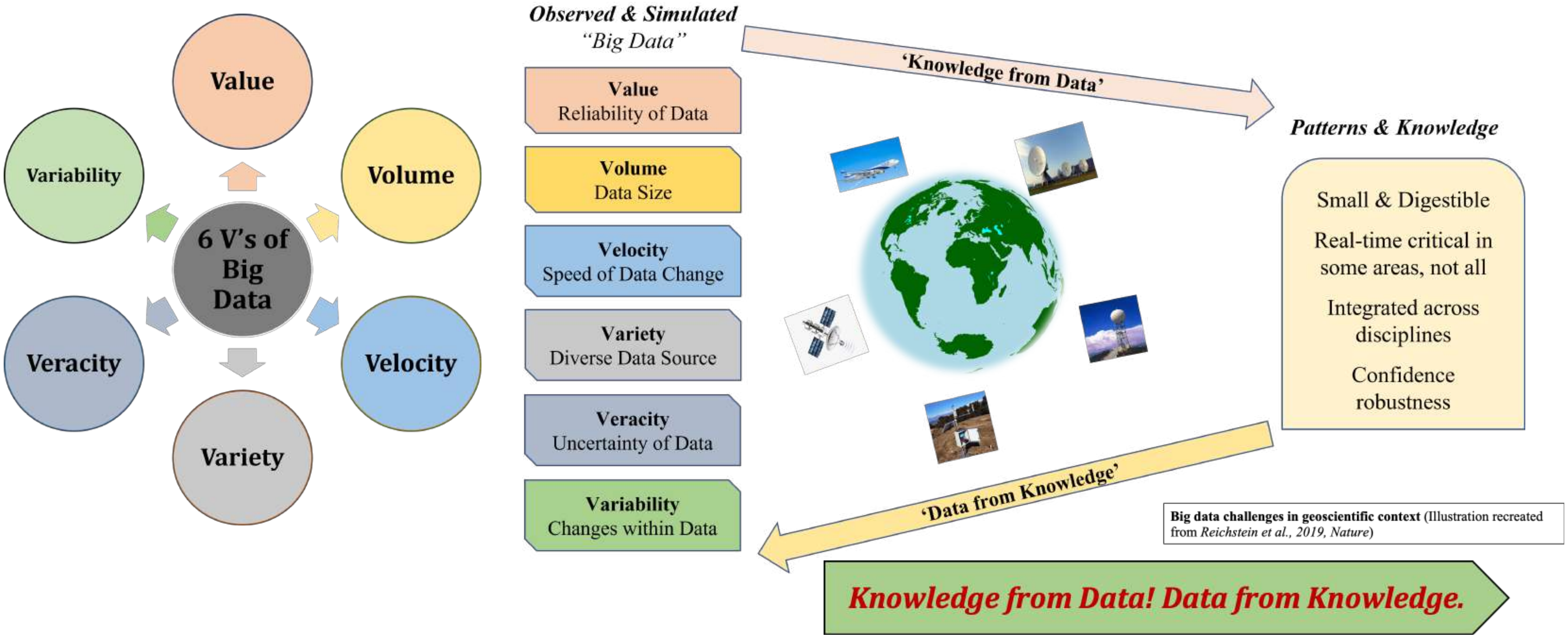




Spatial **biasedness** for intense rainfall, that is, at **95th percentile** for **six RCMs** with respect to the observed precipitation data set from **IMD**

Kumari& Kumar (2023: Intl Journal of Climatology)

Climate data is at ‘Big Data’ scale

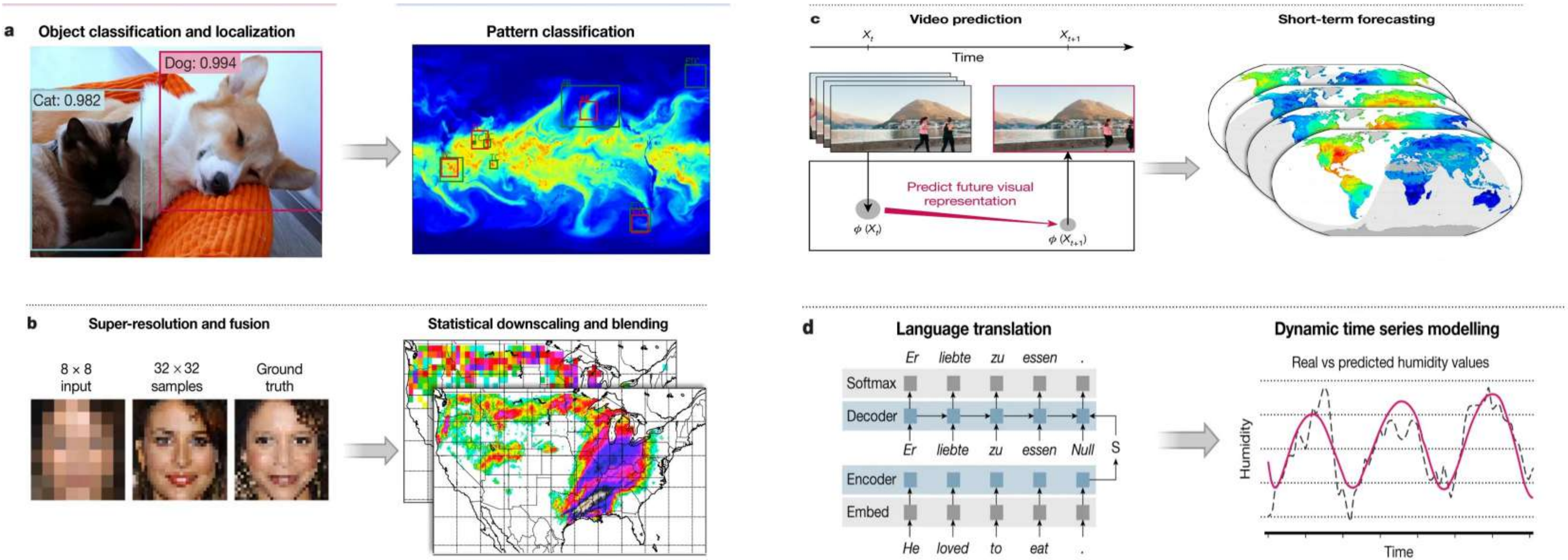


Big data challenges in geoscientific context (Illustration recreated from *Reichstein et al., 2019, Nature*)

- **Climate science is now a ‘Big Data’ problem!**
- **Analysis and interpretation of climate data require the use of more and more, new and regressive techniques.**
- **The current framework of climate modelling can be greatly benefitted from emerging data-driven techniques like Artificial Intelligence/Machine Learning/Deep Learning (AI/ML/DL).**

Deep learning and process understanding for data-driven Earth system science

Machine Learning	Earth Science	Machine Learning	Earth Science



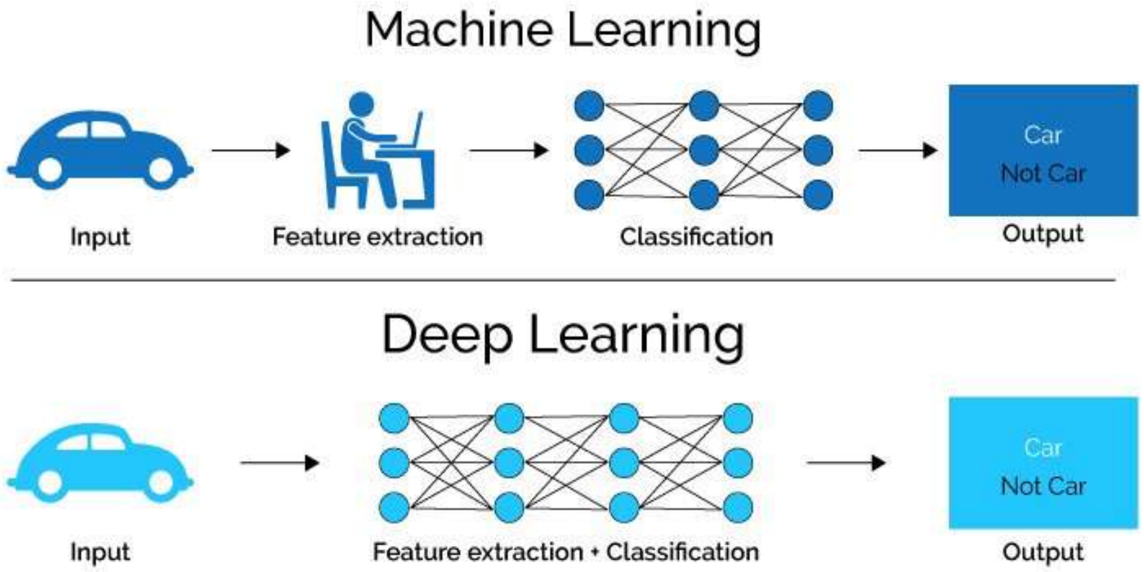
Reichstein et al. (2019)

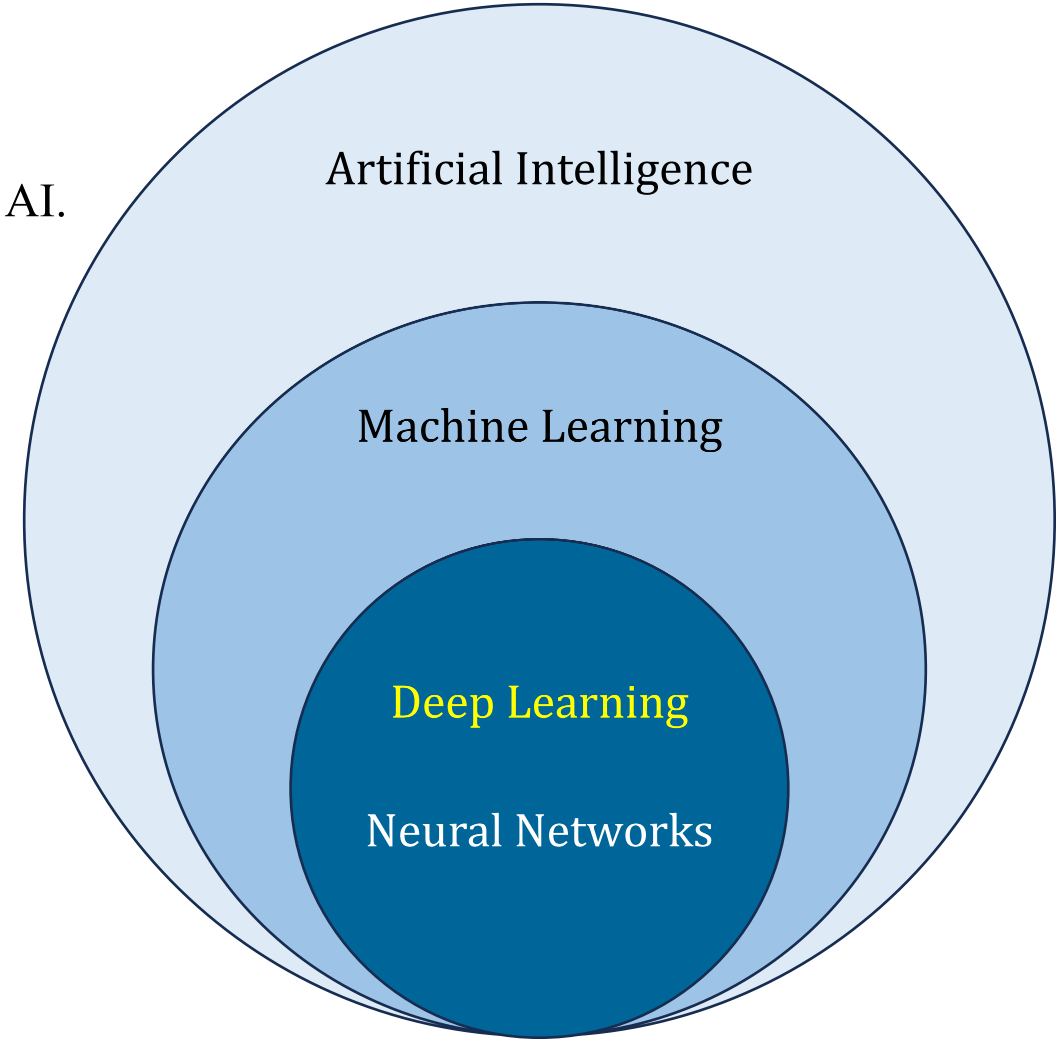
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING (AI-ML)

Artificial Intelligence (AI): Enabling machines to think.
- ✓ Learning, reasoning and problem solving.

Machine Learning (ML): Computer algorithms to accomplish AI.
- ✓ Advanced computational statistics.
- ✓ Training and testing

Deep Learning (DL): Highly advanced ML algorithms.
- ✓ Deep neural networks



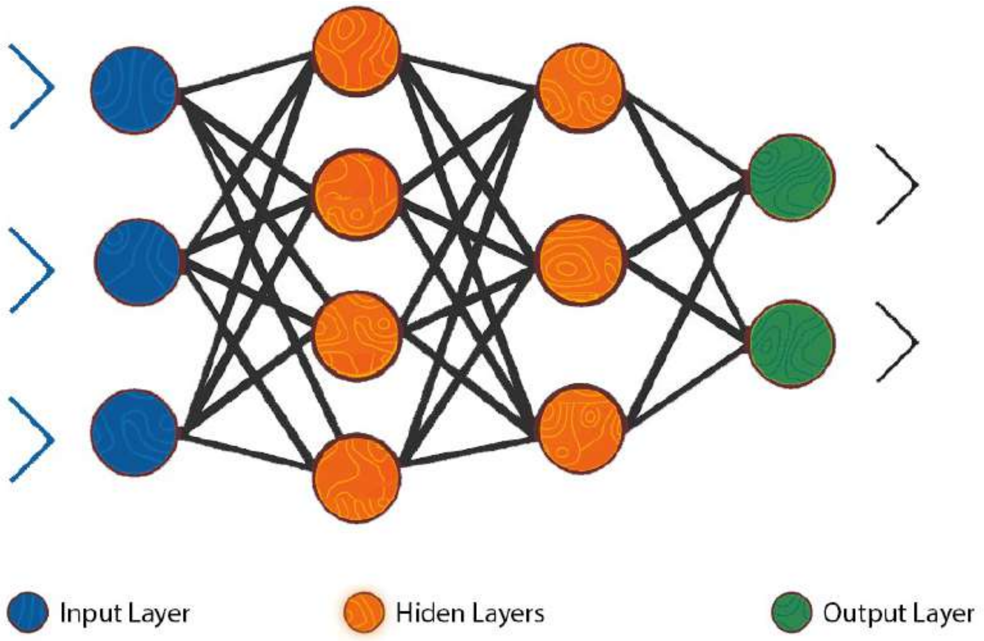


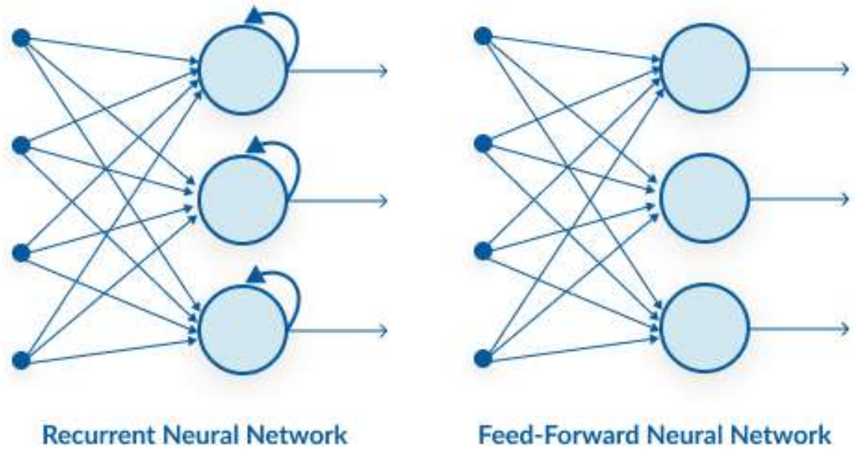
DEEP NEURAL NETWORKS

Artificial Neural Network (ANN):
- ✓ Multiple perceptrons (or neurons)
- ✓ Feed forward neural networks

Recurrent Neural Network (RNN):
- ✓ A looping constraint on the hidden layer of ANN turns to RNN.
- ✓ Long Short-Term Memory (LSTM)

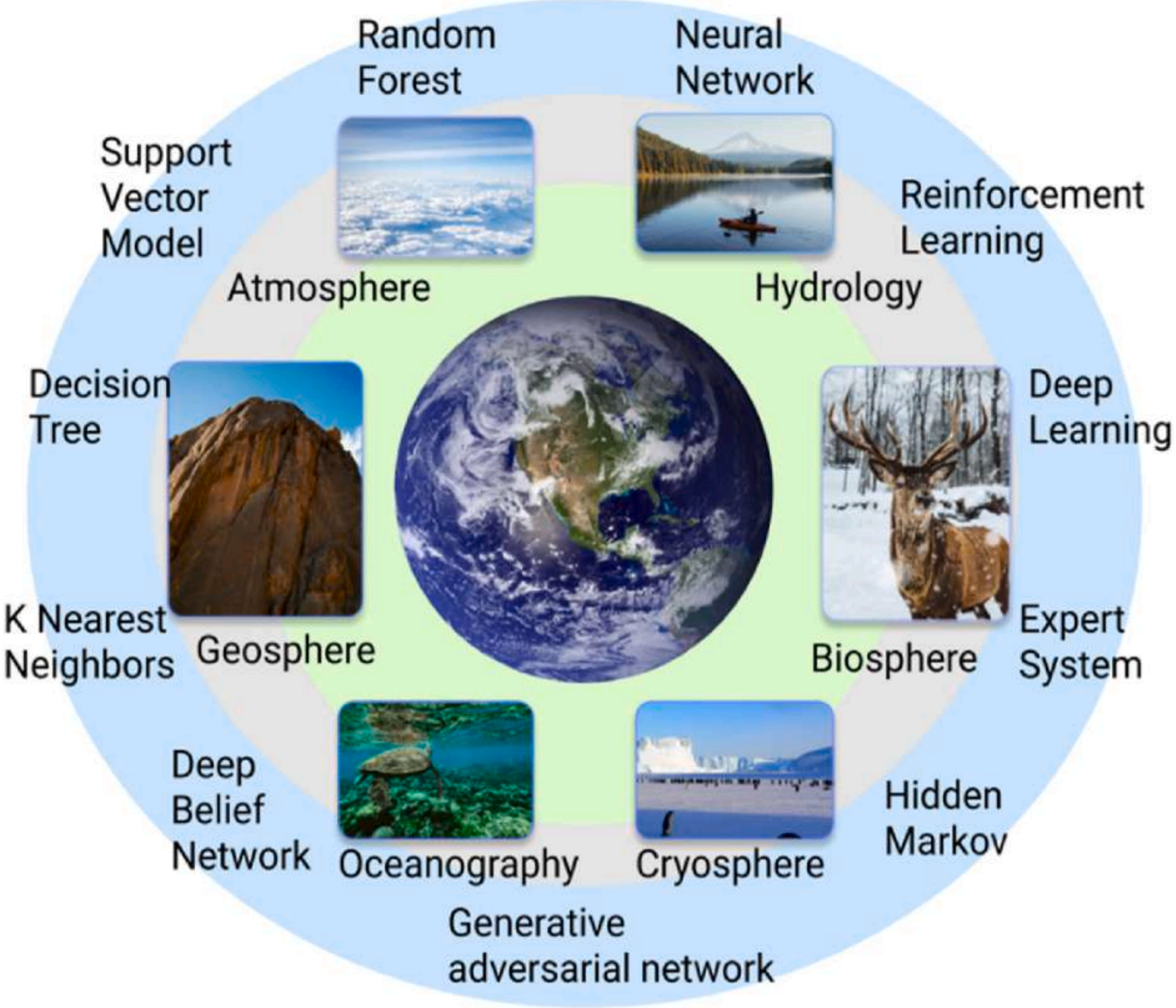
Convolution Neural Network (CNN):
- ✓ Extract the relevant features from the input using the convolution operation.





NEURAL NETWORKS		
ANN	**RNN**	**CNN**
• Tabular data • Image data • Text data	• Time Series data • Text data • Audio data	• Image data • Video processing

Towards Earth Artificial Intelligence



The novel data-driven techniques are capable of identifying patterns from 'big data'.

Sun et al., 2022, A Review on Earth Artificial Intelligence

Different Hybridization to Improve Numerical Weather/Climate Modelling Systems

Earth System Modelling Process flow

Initialization/ DA

- Better ML Retrieval Algorithms
- Fast ML Forward Models
- ML Observation Operators

MODEL

DYNAMICS ⇄ PHYSICS

- Fast ML Radiation
- Fast & Better Microphysics
- New ML Parameterizations
- Fast ML Physics

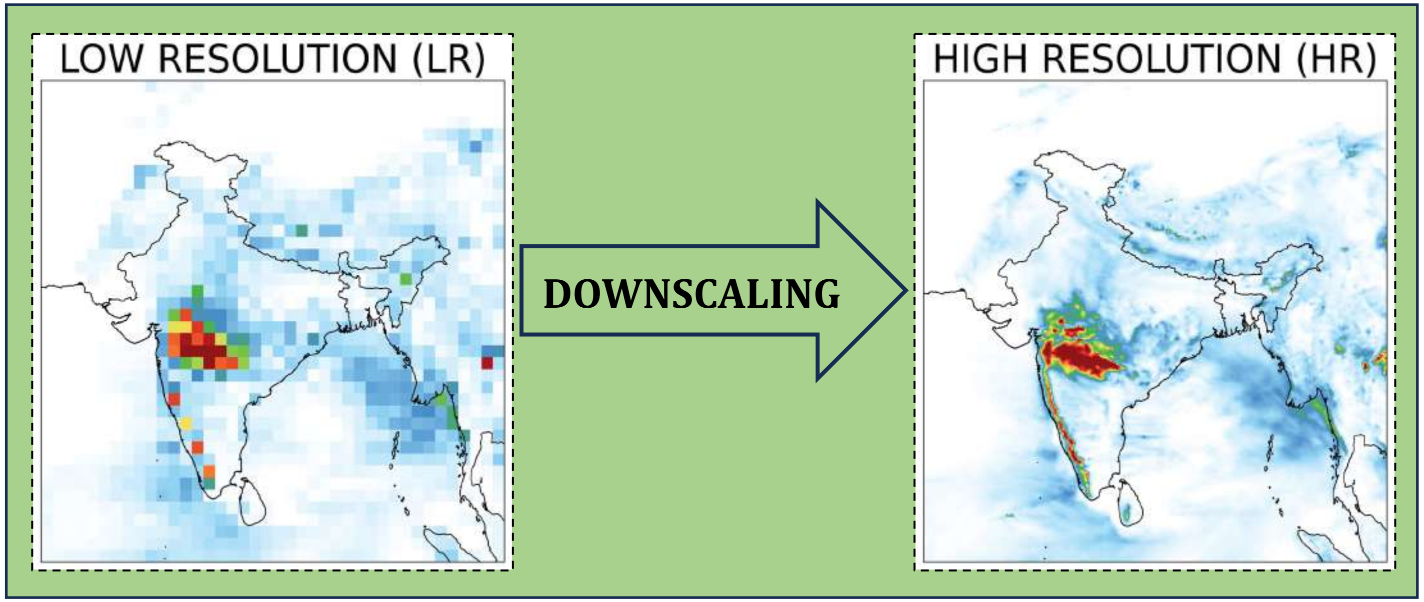
Post-Processing

- Deep downscaling
- ML Bias Corrections
- Nonlinear MOS
- Nonlinear ML ensemble averaging

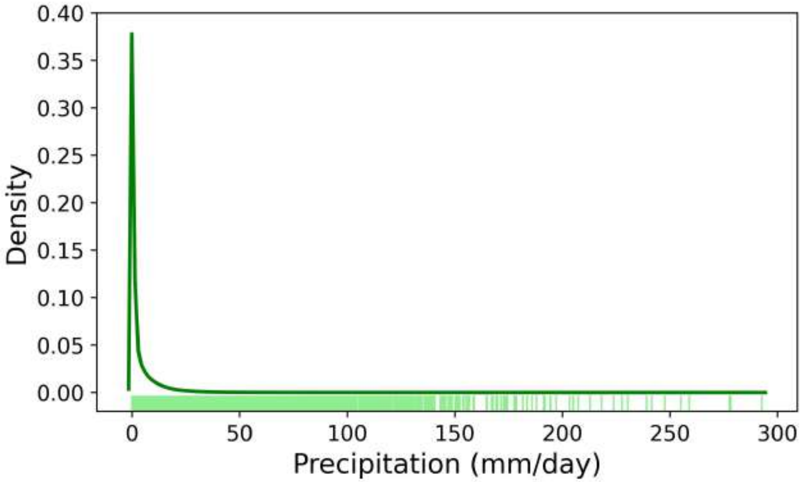
Precipitation Downscaling

High-resolution precipitation data is required by various scientific, societal and management sectors.

- ✓ Climate change assessments
- ✓ Agriculture
- ✓ Hydrological modelling
- ✓ Environmental impact studies
- ✓ Urban planning and infrastructure
- ✓ Disaster management

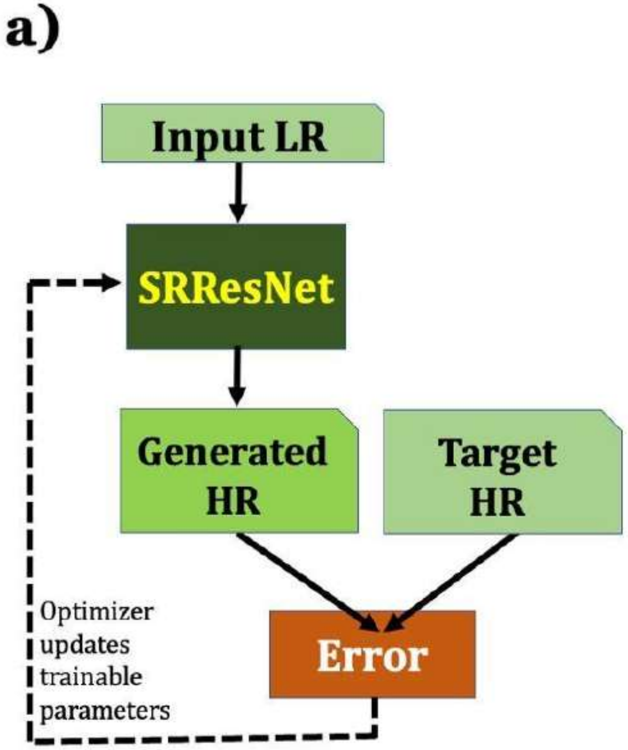


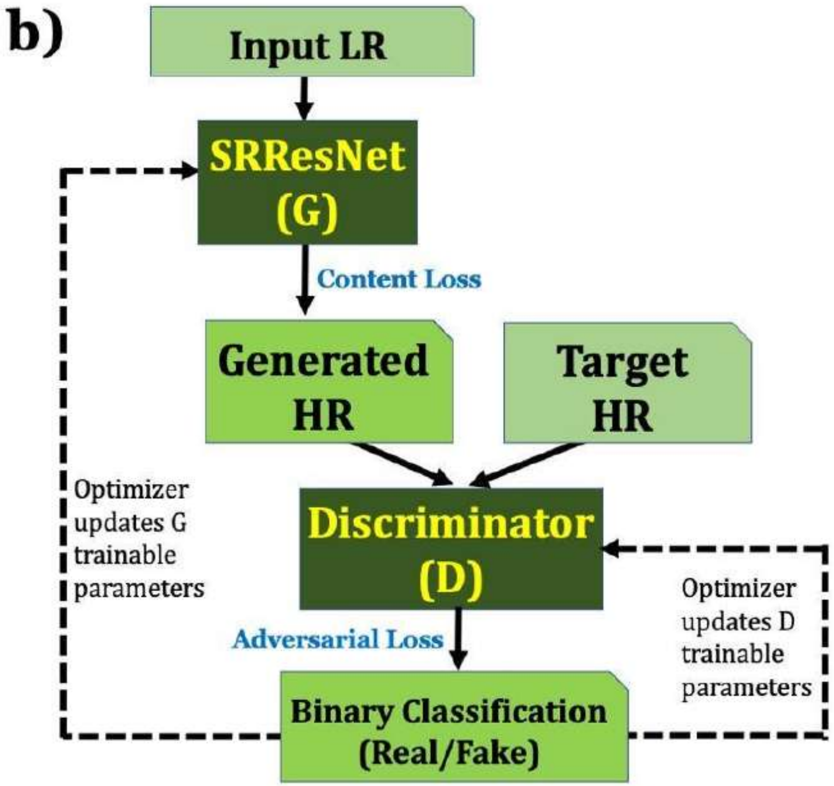
Climate data downscaling
Deriving high-resolution data from a low-resolution Global Climate Models (GCM) dataset.
e.g.: dynamical methods, statistical methods, & AI/ML methods



Precipitation data: ***Skewed distribution with sparse extreme values.***

Generative Adversarial Networks (GANs)



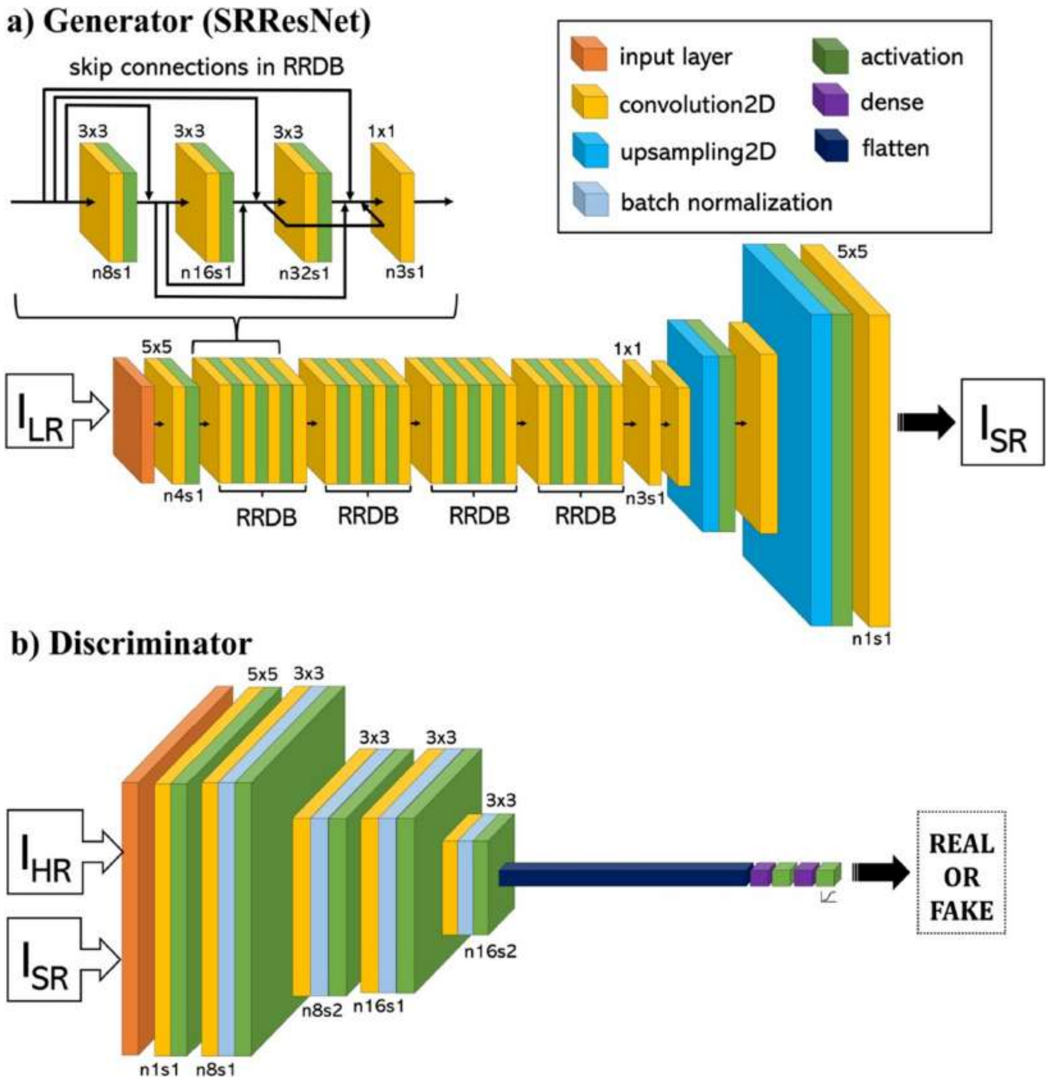


a) Supervised training

b) Adversarial training

- GANs are initially proposed by *Goodfellow et al. (2014).*
- GANs are an advanced class of deep learning algorithms consisting of two separate deep neural networks**: a generator** and **a discriminator.**
- The generator tries to generate fake samples out of a given input/noise, while the discriminator distinguishes the actual samples from the artificial samples generated by the generator.

Super-resolution Generative Adversarial Network (SRGAN)



- SRGAN is optimized by minimizing GAN loss **(perceptual loss)**, which is the weighted sum of content loss and adversarial loss (Ledig et al., 2017).
- **Content loss** is the key to performance of SRGAN, which is minimized at feature space rather than pixel/grid space.

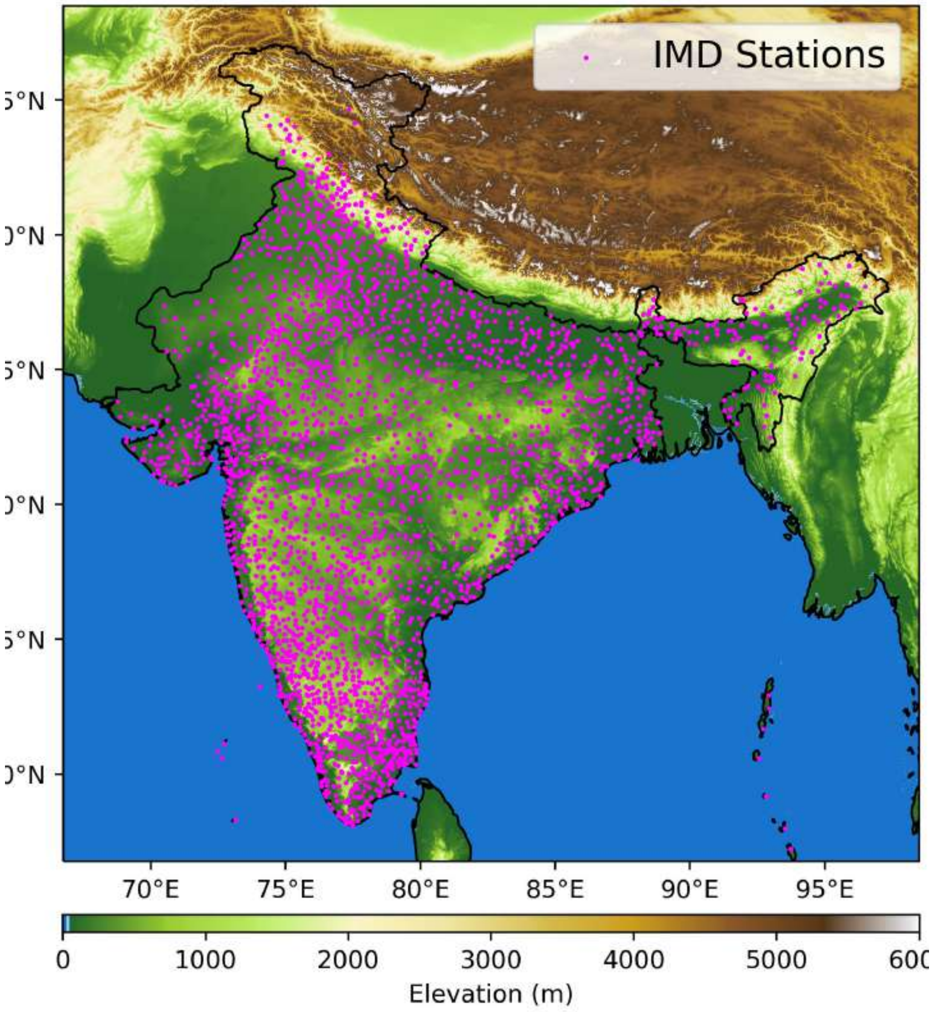
$$l^{SR} = \underbrace{\underbrace{l^{SR}_{X}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

(Ledig et al., 2017)

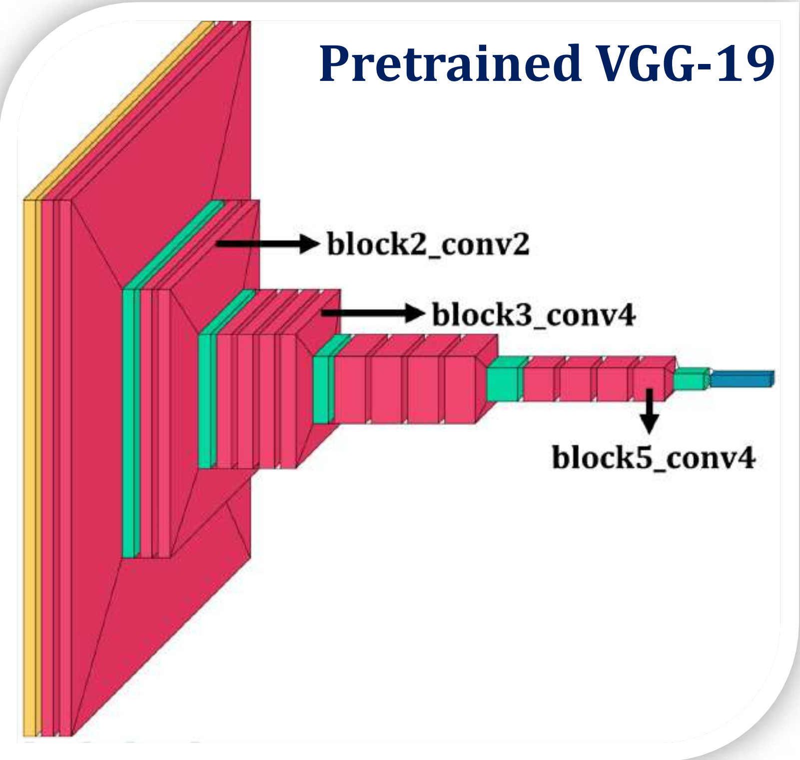
Data & Study Area

- Gridded rainfall data provided by **India Meteorological Department (IMD)**
- IMD provide this data at **low-resolution** **(1ºx1º ~100 KM)** and **high-resolution** **(0.25ºx0.25º ~25 KM)**
- Period of study: **1901-2019**
- In this study, only data for **June-July-August-September (JJAS)** is used.

Training set	Testing set
1901-1999 JJAS	2000-2019 JJAS
(12078 timesteps)	(2440 timesteps)



Experiment Design for Gridded Rainfall



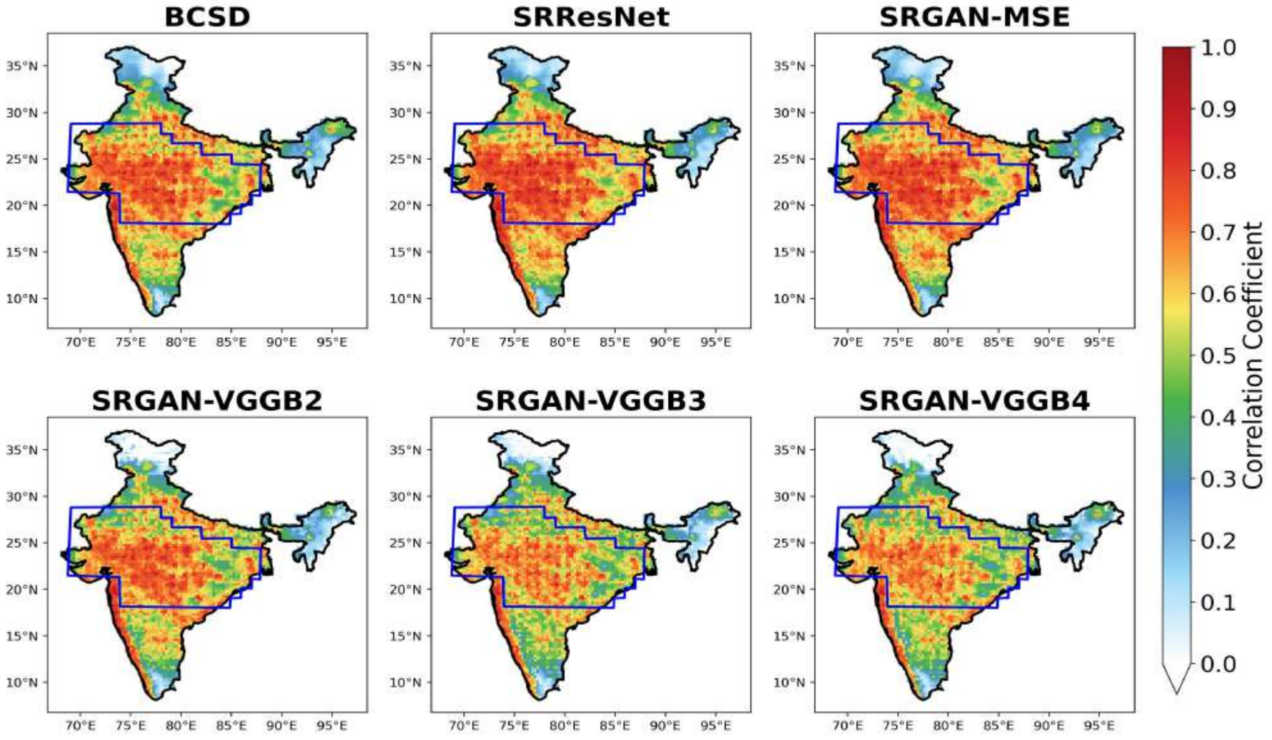
SRGAN Experiments are designed based on pretrained VGG-19 based feature extraction

Experiment (Model)	Description	Hyperparameters & Specifics
BCSD	Bias correction and spatial disaggregation	
SRResNet	The generator network is trained stand-alone in supervised training mode. Optimized by minimizing the MSE loss function.	Learning rate = 0.0001, Batch size = 64, 30% Validation split
SRGAN-MSE	Adversarial training. Content loss is MSE loss between generated and real HR. No feature extraction involved.	Trained on batches of batch size 64, Learning rate = 0.0001
SRGAN-VGGB2	Adversarial training. Content loss optimized at high-level features generated from VGG-19 block2_conv2 layer.	
SRGAN-VGGB3	Adversarial training. Content loss optimized at intermediate-level features generated from VGG-19 block3_conv4 layer	
SRGAN-VGGB4	Adversarial training. Content loss optimized at low-level features generated from VGG-19 block4_conv4 layer	

Results: Overall Agreement

Downscaling Method	CC	RMSE	MAE	PBIAS
BCSD	0.601	13.659	5.834	-3.106
SRResNet	0.657	12.25	5.399	-8.418
SRGAN-MSE	0.647	12.373	5.444	-10.005
SRGAN-VGGB2	0.603	13.28	5.448	-46.169
SRGAN-VGGB3	0.541	14.441	6.245	-23
SRGAN-VGGB4	0.539	14.218	6.086	-29.083

Performance metrics such as RMSE (mm/day), MAE (mm/day), PBIAS (%) and Pearson's correlation coefficient (CC) of the trained models, calculated aggregately over India.

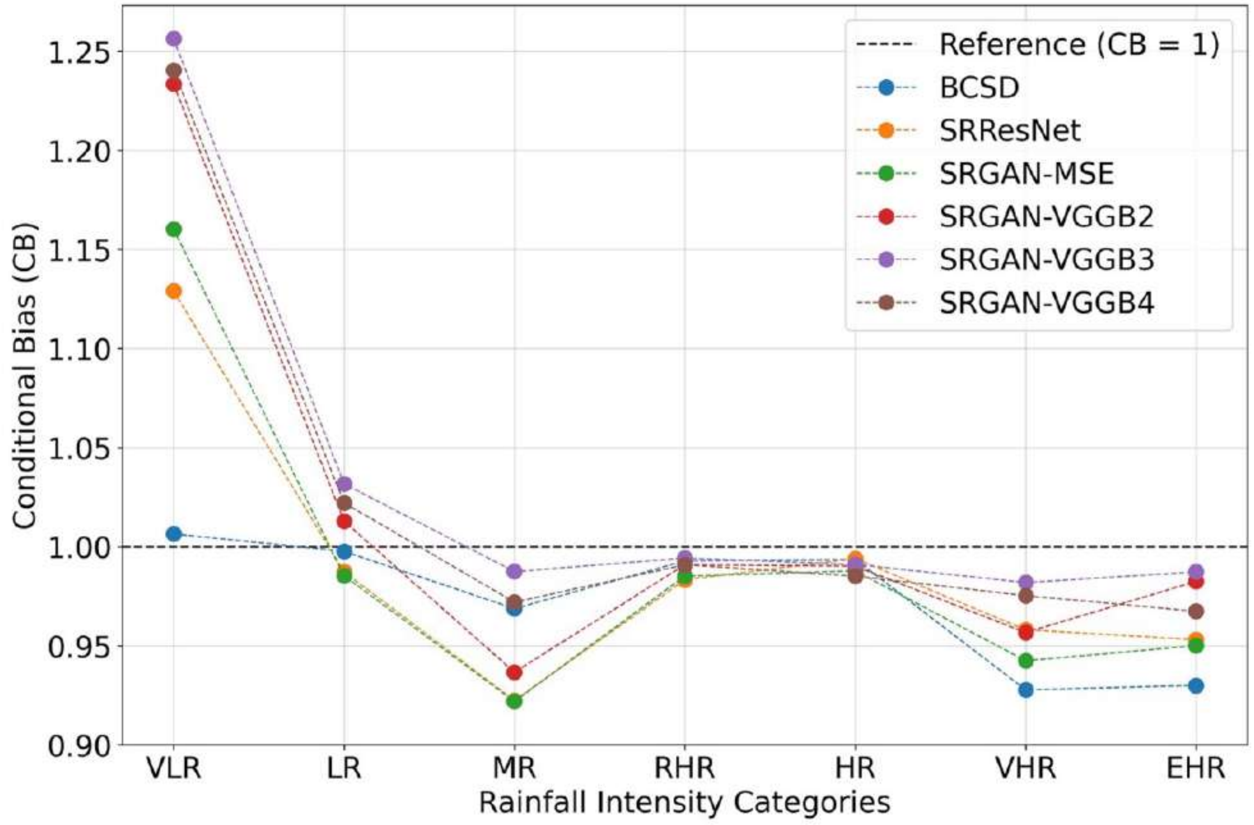


- Spatial map of correlation coefficients (CC) calculated for each grid point of downscaled data from BCSD, SRResNet, SRGAN-MSE, SRGAN-VGGB2, SRGAN-VGGB3 and SRGAN-VGGB4 with IMD-HR during the test period (2000-2021 JJAS).
- The CC values are significant at 95% confidence interval. The region enclosed within the blue bounding box is the Monsoon Core Zone (MCZ).

Results: Conditional Bias

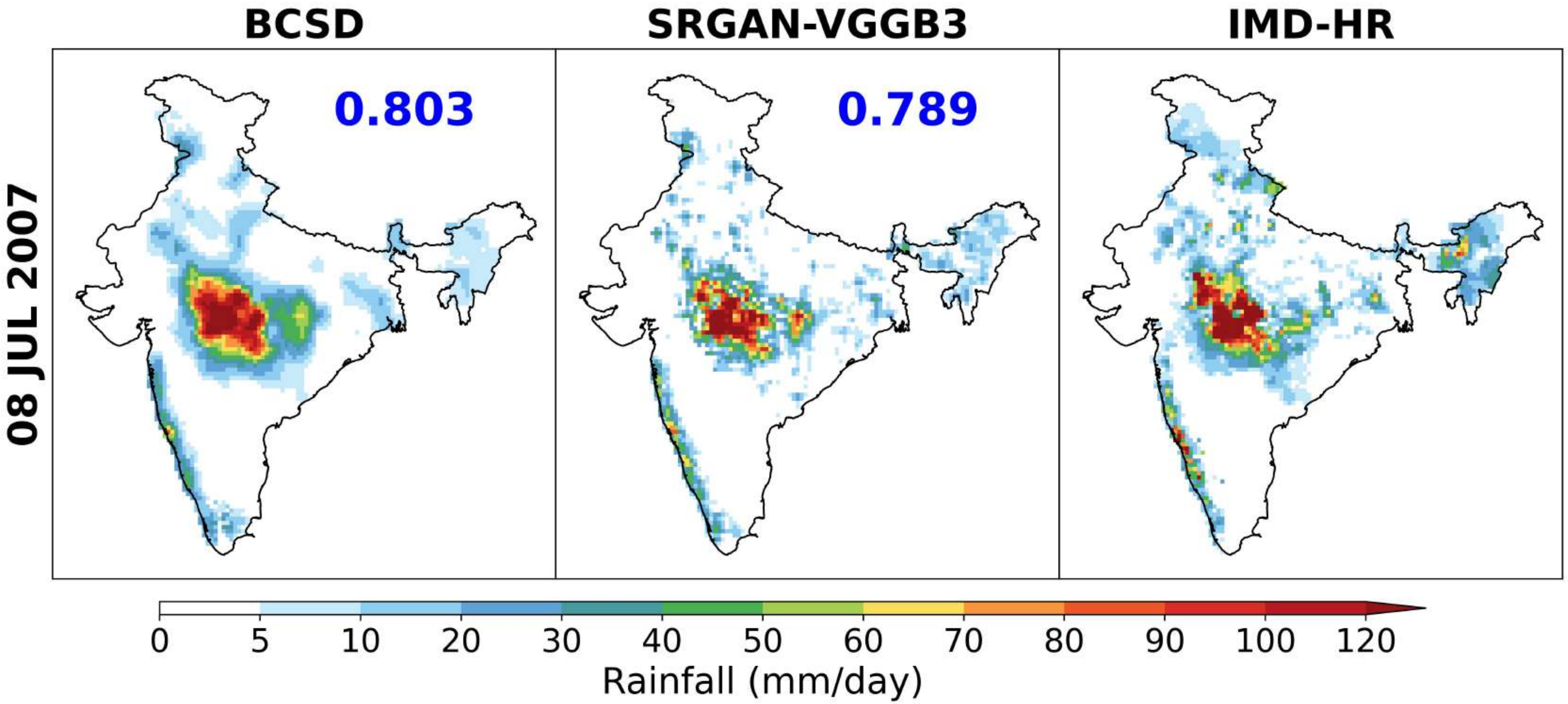
Description term used	Rainfall amount (mm/day)
Very light rain (**VLR**)	0.1 – 2-4
Light rain (**LR**)	2.5 – 7.5
Moderate rain (**MR**)	7.6 – 35.5
Rather heavy rain (**RHR**)	35.6 – 64.4
Heavy rain (**HR**)	64.5 – 124.4
Very heavy rain (**VHR**)	125 – 244.4
Extremely heavy rain (**EHR**)	> 244.5

Various categories of rainfall events and respective rainfall intensity ranges as defined by India Meteorological Department (IMD)



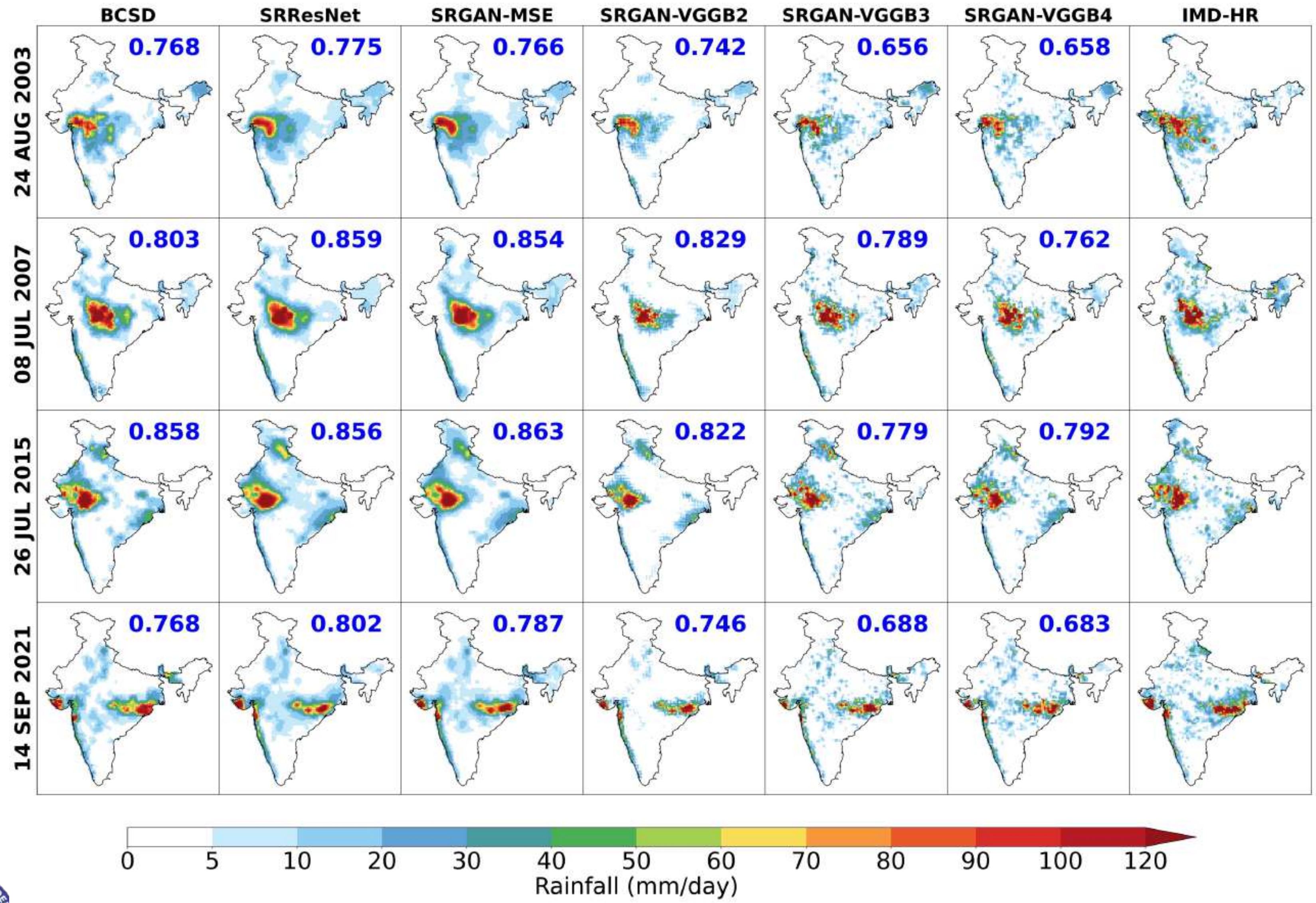
Conditional bias with respect to various Rainfall Intensity Categories as shown by various downscaling methods followed in this study. The description of the rainfall categories are given in Table

Results: A Realtime Examples of extreme precipitation events > 99th Percentile



The **SRGAN** reconstructed a perceptually similar rainfall map compared to **BCSD,** with **better** representation at **sub-grid scale**

Results: Some Realtime Examples of extreme precipitation events > 99th Percentile



Examples showing the actual and generated rainfall maps. Each column indicates the downscaling methods. Each row indicates various samples. The correlation coefficient of each sample with respect to IMD-HR is labelled in blue.

Conclusions

- A **highly efficient image super-resolution technique** (SRGAN) is adopted and applied for downscaling gridded rainfall data over India. The SRGAN is optimized for pre-trained VGG-19-based perceptual loss, and the **SRGAN-VGG-b3c4** variant achieved good perceptual similarity compared to the actual data (from visual inspection).

- It implies that **the GAN training strategy is more powerful than the supervised learning strategy in preserving minute details**. **Also, it is** noticeable that the SRGAN failed to reconstruct the overall statistics of the reconstructed data.

- The complicated deep learning models do not guarantee better results. Finding the optimum balance between model complexity and performance requires more experiments and fine tuning.

Future Outlook

➢ Increase the **physical consistency** of SRGAN model by adding the advantage of static fields (orography and land-sea mask) and dynamic variables (e.g., temperature, humidity, windspeed etc.).

➢ Apply it for **downscaling ESMs/RESMs** future projections.

Thank you for your kind attention!