Safexpress Centre for Data, Learning and Decision Sciences

2nd Annual Workshop on AI/ML Methods in Weather and Climate Modelling

September 12-13, 2025

Day 1: September 12, 2025 **Venue: AC-02-LR-007**

Time	Speaker	Title
2:00 PM - 2:30 PM	Registration	
2:30 PM - 3:20 PM	Krishna AchutaRao, IIT Delhi	Potential applications of AI/ML in climate prediction and climate change attribution
Tea Break / Snacks		
3:50 PM - 4:30 PM	Parthasarathi Mukhopadhayay, IISER Berhampur	Whether the AI/ML model reproduces a better mean state than GCM?
4:30 PM - 5:00 PM	Bedartha Goswami, IISER Pune	A well-calibrated deep learning model for ENSO forecasting
Break		
8:30 PM - 9:20 PM	Aditya Grover, UCLA	Foundation Models for Planetary-Scale Science
Break		
9:30 PM - 10:20 PM	Ignacio Lopez-Gomez, Google Research	Generative downscaling of climate projection ensembles

(Please note: Keynote talks are 50 minutes in duration)

Safexpress Centre for Data, Learning and Decision Sciences

Day 2: September 13, 2025 **Venue: MHP, Admin Building**

Time	Speaker	Title
8:30 AM - 9:00 AM	Registration & tea	
9:00 AM - 9:30 AM	Somak Raychaudhury/Mrutyunjay Mohapatra/Sandeep Juneja	Welcome address by Vice-Chancellor and signing of MoU between IMD and Ashoka University
9:30 AM - 10:30 AM	Mrutyunjay Mohapatra, DGM IMD (chief guest)	Application of AI in weather and climate forecasting
Tea Break / Snacks		

Venue: AC-02-LR-007

Time	Speaker	Title
11:45 AM - 12:45 PM	Sachchida Nand Tripathi, IIT Kanpur	Leveraging AI/ML and Low-Cost Sensors for Data-Driven Air Quality Management: Insights from the Indo-Gangetic Plains
Lunch		
2:00 PM - 2:45 AM	Ravi Nanjundiah, IISc	Using AI for Predicting Atmospheric Parameters
2:45 PM - 3:30 PM	Subimal Ghosh, IIT Bombay	Data, Models, and Decisions: Toward End-to-End Climate Services for Societal Impact
Tea Break / Snacks		
3:55 PM - 4:30 PM	Siddharth Kumar, IITM	Enhancing Weather Forecasts Using a CNN-Based Hybrid Framework for Bias Correction and Ensemble Generation
4:30 PM - 5:00 PM	SCDLDS team (Saptarishi Dhanuka/Atish Kumar)	Modelling monsoon precipitation in India
5:00 PM - 5:30 PM	P. P. Chakrabarti, IIT Kharagpur	Concluding remarks and next steps

(Please note: Keynote talks are 50 minutes in duration)

ASHOKA UNIVERSITY | Safexpress Centre for Data, Learning and Decision Sciences

Abstracts

Title: Potential applications of AI/ML in climate prediction and climate change attribution

Speaker: Krishna Achutarao

Affiliation: IIT Delhi

Title: Data, Models, and Decisions: Toward End-to-End Climate Services for Societal Impact

Speaker: Subimal Ghosh

Affiliation: IIT Bombay

Abstract: A typical climate service involves three essential steps: (i) understanding climate processes, (ii) using this understanding to improve forecasts and predictions, and (iii) engaging with stakeholders to translate forecasts into informed decisions and impact assessments. However, the task is uniquely challenging given the complexity of climate processes, our incomplete scientific understanding, and the uneven availability of multi-source data—leaving vast regions ungauged. The central question is how to design end-to-end, sector-specific decision systems at different scales using the best available science and data.

In this talk, I will first discuss the emerging role of data science—particularly causal analysis—in strengthening our understanding of climate processes and testing their robustness. I will then highlight the application of data-driven methods in global and regional modeling, with special attention to the underrepresentation of human behavior in climate models, which often leads to misinterpretation of critical processes. Finally, I will focus on pathways for designing effective decision-support systems for key sectors such as agriculture, renewable energy, and urban flood management, and reflect on directions for the way forward.

Title: Generative downscaling of climate projection ensembles

Speaker: Ignacio Lopez-Gomez

Affiliation: Google Research

Abstract: Accurate, actionable climate information at kilometer scales is crucial for robust natural hazard risk assessment and infrastructure planning. Simulating climate at these resolutions remains intractable, forcing reliance on downscaling: either physics-based or statistical methods that transform climate simulations from coarse to impact-relevant resolutions. However, physics-based models remain too costly to apply to large climate projection ensembles, and statistical methods often fail to capture compound extremes and crucial spatiotemporal correlations. In contrast, generative artificial intelligence methods such as diffusion models can capture the complex interdependencies of climate processes crucial to risk assessment while remaining computationally scalable. I discuss two ways in which generative artificial intelligence can be used to accelerate progress in regional climate risk assessment: regional climate model emulation, and generative downscaling from observational data. These methods are shown to capture the rich correlational structure of climate data, and enable the assessment of environmental risks such as wildfires, heatwaves, and tropical cyclones. By efficiently translating global projections into actionable local insights, generative AI is poised to become an indispensable tool for climate adaptation.

Title: A well-calibrated deep learning model for ENSO forecasting

Speaker: Bedartha Goswami

Affiliation: IISER Pune

Abstract: Several breakthrough studies in the last five years have shown that deep learning (DL) models can outperform traditional physics-based models in short- to mid-term weather prediction, as well as in seasonal-to-annual predictions of El Nino Southern Oscillation (ENSO) dynamics. However, most existing DL ENSO models are deterministic and the more recent models based on vision transformers are also compute intensive. Here, we propose a probabilistic, well-calibrated, deep learning model of the ENSO that can be more parameter efficient without compromising on performance. Our model outperforms the state-of-the-art and also is uncertainty-aware, which is currently lacking in any of the existing DL ENSO models.

Title: Foundation Models for Planetary-Scale Science

Speaker: Aditya Grover

Affiliation: UCLA Samueli School of Engineering

Abstract : Key global sustainability challenges, ranging from ensuring food and energy security to managing disaster response, critically depend on our ability to accurately forecast weather and project climate. While current approaches are limited by our physical understanding of the atmosphere, improvements in sensory capabilities and large-scale machine learning (ML) present immense opportunities for designing alternative solutions. In this talk, I will discuss key design principles for developing foundation models for atmospheric sciences. Unlike language and vision, scientific domains present unique challenges due to the significant heterogeneity of available datasets. Moreover, downstream tasks in atmospheric sciences require generalizing across a broad range of variables and spatiotemporal resolutions. To address these challenges, I will demonstrate novel strategies for data engineering, model design, spatiotemporal optimization, and cross-modal fine tuning aimed at scalable learning of atmospheric foundation models. Our resulting models exhibit exceptional skill, speed, and adaptability across a range of predictive tasks in weather and climate science. Towards the end, I will also announce IndiaWeatherBench, a curated dataset and benchmark to advance AI-driven weather modeling for India.

Safexpress Centre for Data, Learning and Decision Sciences

Title: Enhancing Weather Forecasts Using a CNN-Based Hybrid Framework for Bias Correction and Ensemble Generation.

Speaker: Siddharth Kumar

Affiliation: Indian Institute of Tropical Meteorology

Abstract: Short-to-medium-range rainfall prediction has improved significantly in recent years. Weather prediction models such as GFS, GEFS, and the newly launched Bharat Forecast System (BFS) are performing well in both qualitative and quantitative terms. However, there remains scope for improving the skill and spatial structure of these forecasts. Ensemble prediction systems are effective in representing forecast uncertainty. For very high-resolution models, however, running multiple simulations and generating ensembles is computationally challenging.Emerging data-driven methods, particularly neural networks, offer promising alternatives for rainfall prediction and bias correction. In this study, we employed Convolutional Neural Networks (CNNs) for bias correction and ensemble generation of rainfall forecasts. Meteorological parameters directly linked to convective systems were incorporated as input features, while observed rainfall was used as the target variable. For the analysis, ERA5 reanalysis rainfall data (0.25°) and GFS rainfall data were used as raw input, and IMD rainfall data (0.25°) served as the target for training and evaluation. Our findings indicate that CNN-based models combined with ensembles significantly improve forecast skill and spatial structure. The evaluation metrics—spatial correlation, root mean square error (RMSE) and mean absolute error (MAE)—demonstrated significant improvement compared to the raw rainfall outputs from model and reanalysis. This approach demonstrates strong potential to enhance overall rainfall forecasts, and a system based on this methodology is currently under development for eventual operational implementation.

Title : Using AI for Predicting Atmospheric Parameters

Name: Ravi S Nanjundiah

Affiliation: Centre for Atmospheric and Oceanic Sciences, and Divecha Centre for Climate Change, Indian Institute of Science

Abstract: Predicting atmospheric parameters at various scales is a very challenging task for both numerical and AI based models. In this talk, I will discuss predicting various parameters such as rainfall and temperature at medium range and seasonal scales using various AI techniques. I will also discuss about some of our ongoing work on using the latest AI weather forecast models and leveraging them for heat stress studies with particular reference to Karnataka.

Title: Leveraging AI/ML and Low-Cost Sensors for Data-Driven Air Quality Management: Insights from the Indo-Gangetic Plains

Speaker: Sachchida Nand Tripathi

Affiliation: IIT Kanpur

Abstract: In India, the sparse air quality monitoring infrastructure beyond urban centers limits our understanding of fine particulate matter (PM2.5) spatiotemporal variability, exposure dynamics, and their health and ecological impacts. To address this challenge, the Ambient Air Quality Monitoring in Rural Areas using Indigenous Technology (AMRIT) project was launched, deploying 1,400 low-cost air quality sensors (LCS) across diverse demographics and land-use settings in the states of Uttar Pradesh and Bihar, within the Indo-Gangetic plains (IGP). Additionally, the Dynamic Hyper-local Source Apportionment (DHSA) project was initiated to explore the potential of LCS for real-time source apportionment, facilitating pollution source identification and providing policymakers with critical data for effective air pollution control. Together, these initiatives aim to capture regional pollution dynamics and impacts beyond urban centres for the first time in India.

The AMRIT network effectively captured fine-scale pollution patterns across Bihar, capturing distinct urban–rural gradients in pollution exposure. Machine learning–based exposure assessment revealed strong spatiotemporal heterogeneity, with persistently elevated levels in north-western districts and a notable associated mortality burden. Data-driven analysis identified multiple distinct air-sheds on both sides of the Ganges, reflecting similar exposure trends, with the north-western and north-central regions being most affected. The LCS network also detected pollution hotspots by integrating satellite-driven deep learning across different temporal scales.

Under DHSA, machine learning methods enabled highly accurate real-time source apportionment, demonstrating the potential of LCS networks for detailed, spatiotemporal source tracking. These approaches establish a foundation for decentralized mitigation strategies and evidence-based policy, providing a cost-effective alternative to conventional monitoring and a scalable solution for data-sparse regions worldwide.

Building on AMRIT and DHSA, a Decision Support System (DSS) integrating LCS data, satellite observations, and AI/ML can transform IGP air quality management. The DSS would deliver real-time, spatially resolved insights on pollution levels, exposures, and sources; identify hotspots; and guide interventions at village, district, or air-shed levels. Air-shed delineation with dynamic source attribution enables locally tailored strategies, while scenario analysis can evaluate policies such as crop residue management or emission controls. By translating complex data into actionable intelligence, the DSS empowers decentralized decision-making, strengthens policy frameworks, and supports progress toward improved air quality, reduced health burdens, and sustainable development goals.

Safexpress Centre for Data, Learning and Decision Sciences

Title: Whether the AI/ML model reproduces a better mean state than GCM?

Speaker: Parthasarathi Mukhopadhaya
Affiliation: IISER Odisha

Abstract :

While the Global Climate Models/General Circulation Models (GCMs) are approaching convection permitting resolution (BFS, IFS, ICON etc.) attempting to improve the mean state of the global climate and it's variability, the advancement of AI/ML based models showed remarkable advancement in the recent years particularly performing equivalently with the ECMWF IFS in weather scale. However there are challenges. While the convection permitting or cloud resolving GCMs need huge computing power as against the AI based model, the extreme events simulation by AI/ML models appear to resolve lesser variability. However, what remains elusive in the midst of promising advancement of AI models is, how good is the mean state simulation in an AI model (particularly in the backdrop where it is known that climate models do not have good fidelity in capturing the global mean precipitation etc.)? To answer this important question, I would show some results from hindcast of Graphcast model, the simulation of mean state and it's variability over the whole globe.