Abstract

India receives 78% of its annual aggregate rainfall in the summer monsoon season (June, July, August and September), which has an important impact on the country's society and economy. The seasonal variations of Indian Summer Monsoon Rainfall (ISMR) are influenced by global climatic conditions. The long (beyond 2 years) and short lead (less than 12 months) prediction of ISMR is important for policy-making and disaster management. In recent years, ISMR prediction has achieved a high accuracy by employing various numerical (involving numerical equations) and statistical (involving statistical prediction models) models . Recent advances in Artificial Intelligence (AI) are being used to improve short-lead and long-lead predictions of ISMR, particularly in relation to global climate variability and change. Furthermore, recent developments in Explainable AI (XAI) allow us to understand the workings of such predictive models, which in turn may provide important knowledge about the influence of different variables on ISMR. With these motivations, this thesis aims to understand the relationship of ISMR with global climate, with the purpose of forecasting and simulation using state-of-the-art AI models. We explore causality-based methods such as Granger Causality and PCMCI to understand the nature of the relationship of ISMR with global climate oscillations such as El Niño Southern Oscillation (ENSO), Atlantic Multidecadal Oscillation (AMO). Statistical approaches for feature selection are used to identify the best drivers that can improve the long-lead predictability of ISMR. We show that such analysis can also help to identify the drivers of past annual extremes of ISMR. Long lead (2 years prior) ISMR forecasting using Convolutional Neural Network (CNN) with global maps of geophysical variables has been able to achieve almost 87 % accuracy. AI models, especially Deep Neural Networks (DNN) require huge amounts of reliable data for successful training, whereas the observational data is often not sufficient. This necessitates the use of simulated data from Earth System Models (ESM), which are used to simulate weather and climate by solving complex physics- and chemistry-based equations at different spatio-temporal scales. However, they suffer from computational bottlenecks and require massive computing resources, and also from biases arising from model design. This thesis also aims to address this issue by developing a Generative AI model that can emulate global Sea Surface Temperature (SST), thereby providing a data-driven surrogate for process-based ESMs. We train these models with ESM simulations, and then fine-tune them with observational data to partially correct the biases. In recent years, various generative models like Generative Adversarial Network (GAN), Variational Auto Encoder (VAE), and Diffusion models have been used in various scientific domains to generate realistic data. In this research, several generative models are compared for emulating the ESMs to generate realistic climate data, especially global SST. The results demonstrate that it is possible to develop an accurate and simple climate emulator using Pix2Pix GANs. We develop CV-GAN, which is capable of simulating the seasonal and decadal variability of global SST for a period of 1000 years. The thesis also indicates the potential of generative AI-based emulator models to become a foundation for ESMs.

Keywords: Granger Causality, Generative AI, Sea Surface Temperature, ISMR, Explainable AI, Climate Emulator.