Report for Independent Studies

Soumyajit Saha

In the course of the independent studies under Dr. Zhaoxia Pu, my work has mainly focused on studying and implementing different machine learning techniques that can be used in Numerical Weather Prediction (NWP). The basic concept of NWP is to solve a set of partial differential equations (PDEs) that govern atmospheric motion and evolution. The complete set of seven equations with seven unknowns which governs the evolution of the atmosphere include Newton's second law or conservation of momentum (three equations for the three velocity components), the continuity equation or conservation of mass, the equation of state for ideal gases, the first law of thermodynamics or conservation of energy, and a conservation equation for water mass. I also got some insight about different NWP models like Finite Differential Equations, Spectral models, Grid staggering models, and boundary conditions. This information was obtained from the book chapter "Numerical Weather Prediction Basics: Models, Numerical Methods, and Data Assimilation". Land-Surface and Ocean Models and Coupled Numerical Models in NWP were also discussed. The weather forecasts produced by operational centres use data assimilation to estimate initial conditions for the forecast model from observations. Data assimilation is referred to a sequential time-stepping method, in which a previous model forecast is compared with newly received observations, which leads to the model state being updated in order to show the observations, and it initiates a new forecast.

The next phase of my work, I mainly focused on doing a literature survey on machine learning or deep learning techniques used in NWP, weather forecasting, parameterization and data assimilation. These studies gave an insight to different ML/DL techniques like Linear Regression, Support Vector Machines, Artificial Neural Network, U-Net based Deep Neural Network, Random Forest regression, Recurrent Neural Network, Generative Adversarial Network (GAN), Convolutional Neural Network, etc. and how they can be used for NWP, weather forecasting, parameterization and data assimilation. As per the guidance of Professor, I narrowed down by focus to ML/DL techniques involved in parameterization and data assimilation techniques in Lorenz 96 Model.

The Lorenz 96 model is a dynamical system formulated by Edward Lorenz in 1996. In its simplest version the model is described by a periodic system of K $(k = 1, \dots, K)$ ODEs:

$$\frac{dX_k}{dt} = \underbrace{-X_{k-1}(X_{k-2} - X_{k+1})}_{\text{Advection}} \underbrace{-X_k}_{\text{Diffusion}} \underbrace{+F}_{\text{Forcing}}$$

The first term on the right hand side is an advection term, while the second term represents damping. F represents an external forcing term. Here, the variables X_k can be interpreted as values of some atmospheric quantity (e.g., temperature, pressure or vorticity) measured along a circle of constant latitude of the earth. The latitude circle is divided into K equal sectors, with a distinct variable X_k for each sector such that the index $k = 1, \dots, K$ indicates the longitude. For parameterization, a two-level version of Lorenz 96 is used. For this another periodic variable Y with its own set of ODEs is added. The X and Y ODEs are linked through coupling term which is the last term in both equations below. Each X has j Y variables associated with it.

$$\frac{dX_k}{dt} = \underbrace{-X_{k-1}(X_{k-2} - X_{k+1})}_{\text{Advection}} \underbrace{-X_k}_{\text{Diffusion}} \underbrace{+F}_{\text{Forcing}} \underbrace{-hc\bar{Y}_k}_{\text{Coupling}}$$

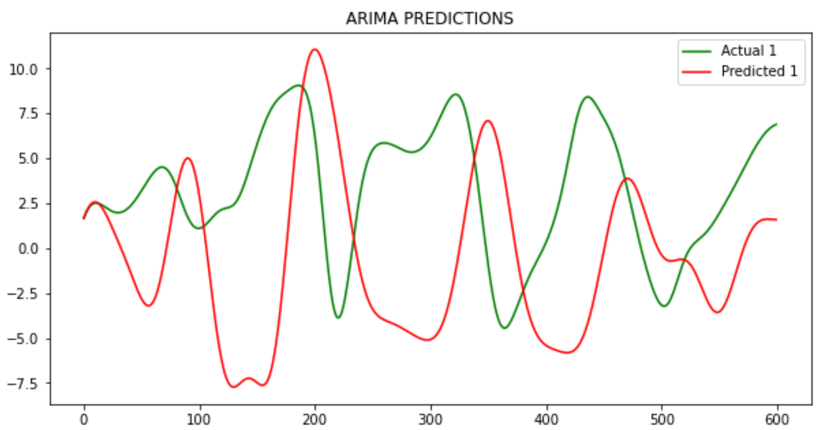
$$\frac{1}{c}\frac{dY_{j,k}}{dt} = \underbrace{-bY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k})}_{\text{Advection}} \underbrace{-Y_{j,k}}_{\text{Diffusion}} \underbrace{+\frac{h}{J}X_k}_{\text{Coupling}}$$

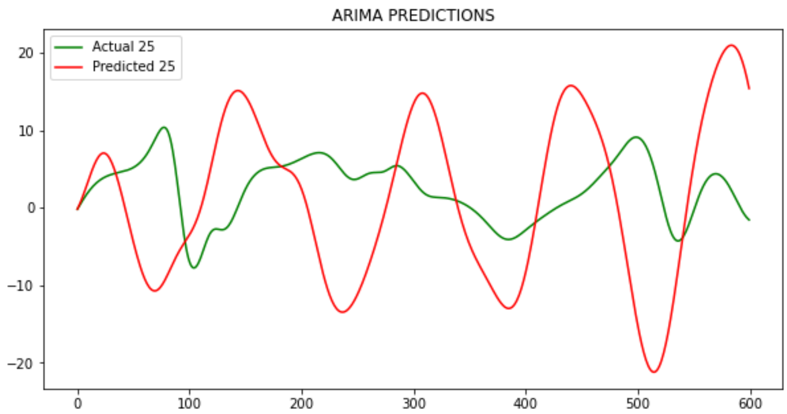
In the next phase, I read two papers which used ML/DL techniques for performing parameterization in Lorenz 96 model. The papers were titled "Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model", and "Data-Driven Super-Parameterization Using Deep Learning: Experimentation with Multiscale Lorenz 96 Systems and Transfer Learning". I got an idea on how to preprocess data for predicting the parameterization using ML/DL techniques and shared a summary of the techniques used in both the papers. The main reason behind the usage of ML/DL techniques for performing parameterization is to avoid the computational overhead associated with generating data using integration techniques. In both of the papers, some data points for the parameter term were initially generated using RK-4 method, which was further split into training and testing data for performing the prediction.

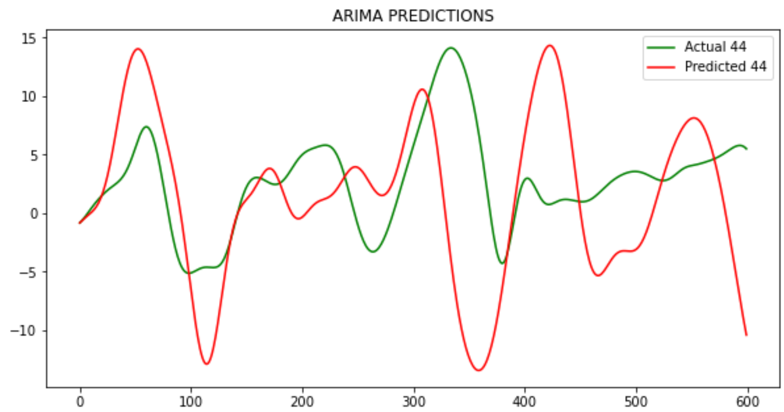
After getting the ideas of the experiment performed, now I tried to apply ML techniques to predict future data points for one unknown variable X of the Lorenz 96 model. At first, I tried to get 3000 data points using RK-4 method on the ODE:

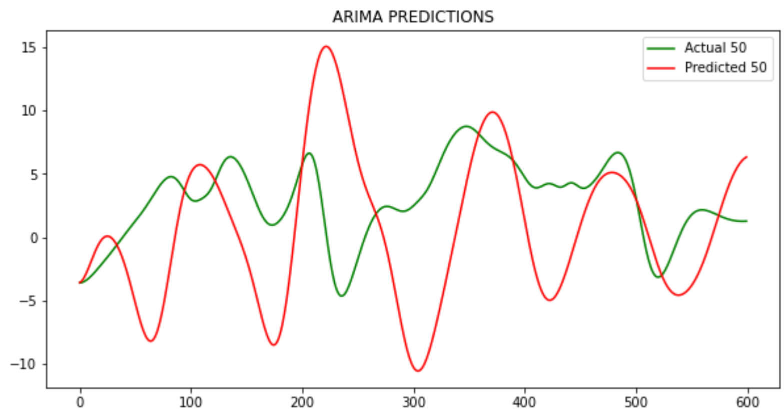
$$\frac{dX_k}{dt} = \underbrace{-X_{k-1}(X_{k-2} - X_{k+1})}_{\text{Advection}} \underbrace{-X_k}_{\text{Diffusion}} \underbrace{+F}_{\text{Forcing}}$$

It was done with 50 variables i.e. $K = 50$ and $F = 8$, for setting the initial state. I set the time range from 0 to 30 with time step of 0.01. I split these dataset into 80: 20 ratios for training and testing. First, I tried to experiment with Autoregressive Integrated Moving Average (ARIMA) model for prediction of the future data points. The average Root Mean Square Error (RMSE) over 50 variables for the test dataset was 81.25857586840742. Also, when I increased the size of the dataset to 30000, it was observed that the predicted data point values are getting converged to 0 base line. Given below are some plots for some selected variables to show the prediction of data points with ARIMA model for the test datasets.

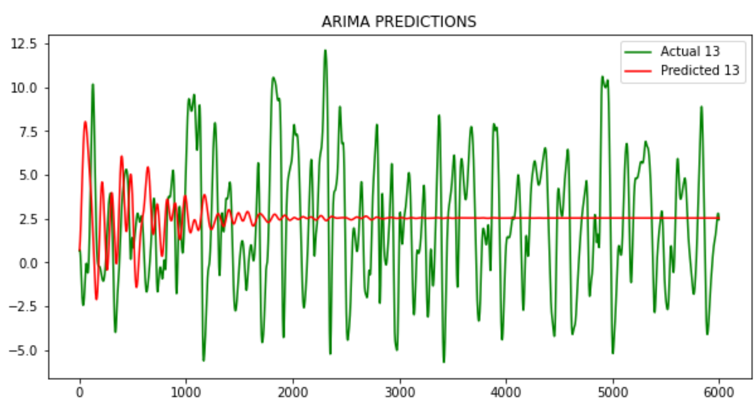


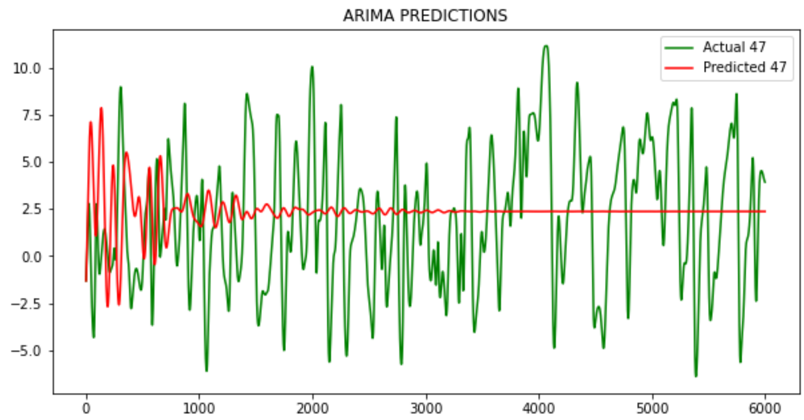






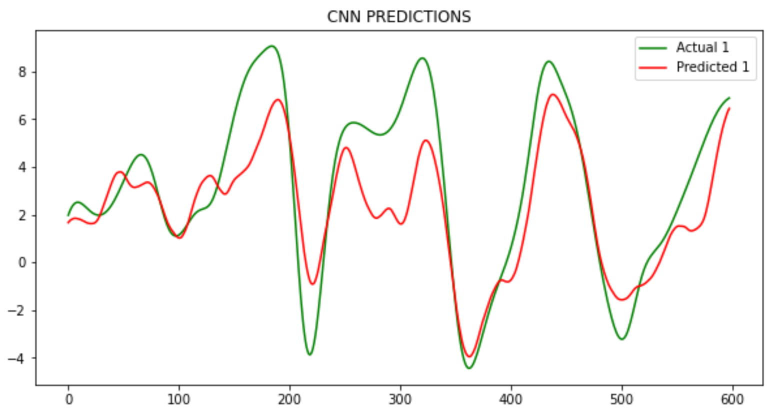
And when I increased the dataset size to 30000, the plot generated for the predictions of the test dataset is given below for 2 variables.

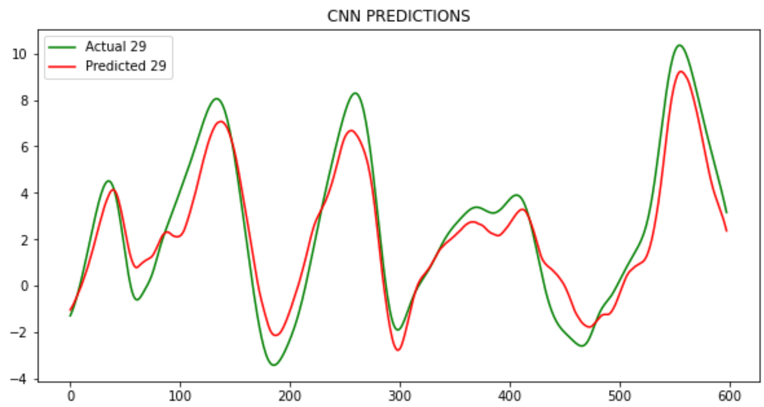


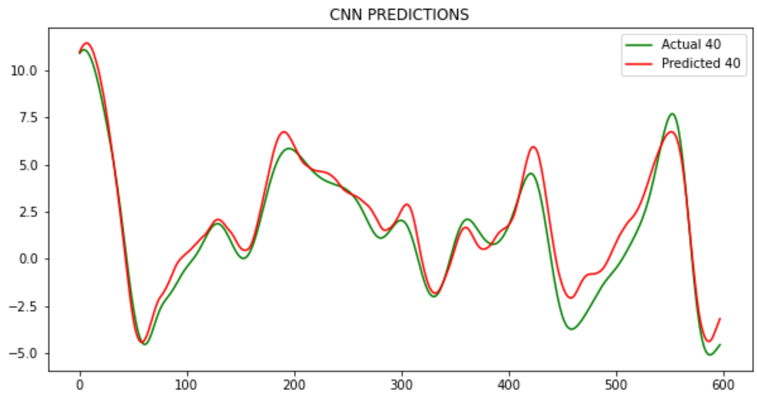


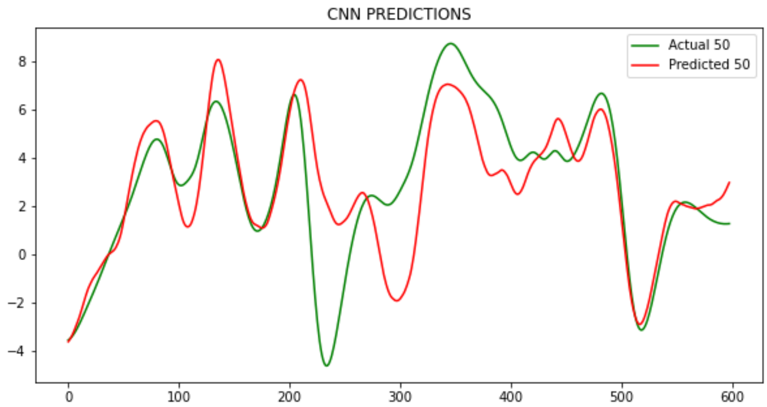
So, clearly it suggested the ARIMA model is not efficient in predicting the data points for Lorenz 96 model taking into consideration the RMSE values and the convergence of the plot to 0 base line for high dimensional datasets.

I then tried using a Convolutional Neural Network (CNN) model for the prediction using the same 3000 data points. The preprocessing of the data points was a bit different from the above approach. For this technique, I have set $n = 10$ and while iterating over the entire data points, $n - 1$ data points were put to X and n^{th} data point was put to Y label. It was done for the all $n = 10$ length window over the dataset. The CNN model consisted of a Convolution 1D layer with $filters = 32, kernel\_size = 3, activation = 'relu', input\_shape = (n, K, 1)$, followed by flatten and Dense layer. The loss was computed using RMSE and optimizer was set to 'adam'. For training the dataset, epochs were set to 50. The dataset was divided into same 80: 20 ratios for training and testing. The average RMSE value over 50 variables for test data was 1.3914003584646872. Given below are the plots for predicted data points using CNN over the test data points using CNN for some variables.









It can be clearly observed that CNN outperformed ARIMA model by a high margin and predicted the data points efficiently with a low RMSE. It can also be observed that the RMSE's were highest for the edge variables in CNN predictions.

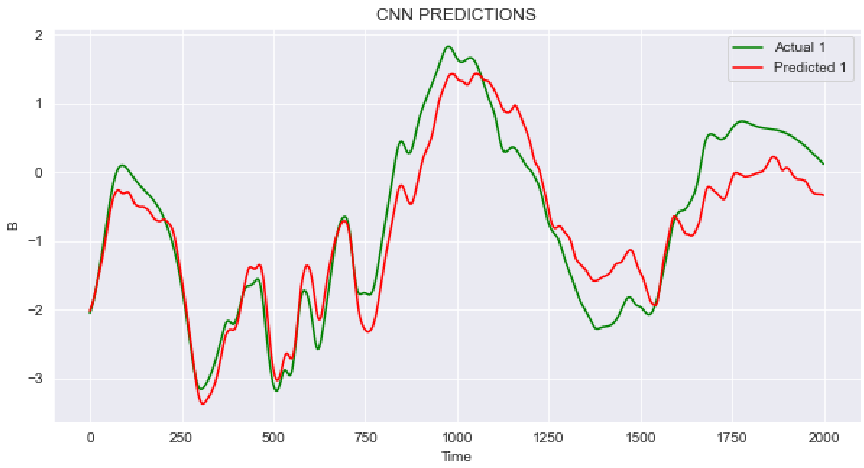
In the next phase of the experiment, I focused on prediction on 2 Two Level Lorenz 96 model with 2 unknown variables X and Y. The ODEs are described as follows:

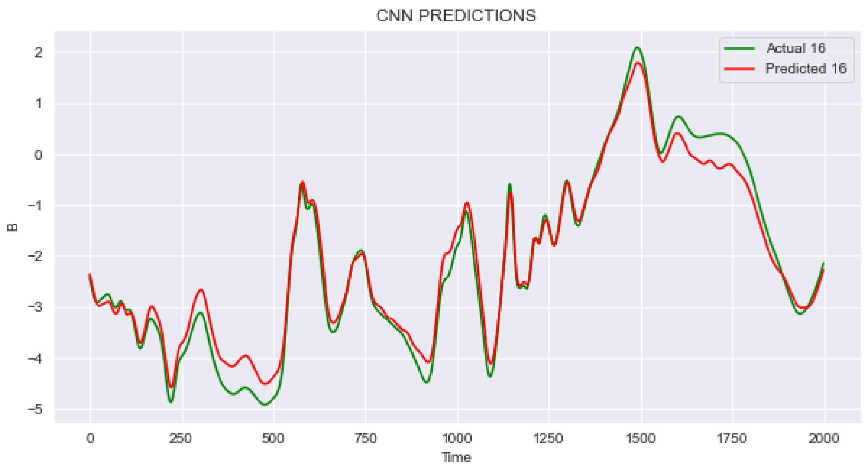
$$\frac{dX_k}{dt} = \underbrace{-X_{k-1}(X_{k-2} - X_{k+1})}_{\text{Advection}} \underbrace{-X_k}_{\text{Diffusion}} \underbrace{+F}_{\text{Forcing}} \underbrace{-hc\bar{Y}_k}_{\text{Coupling}}$$

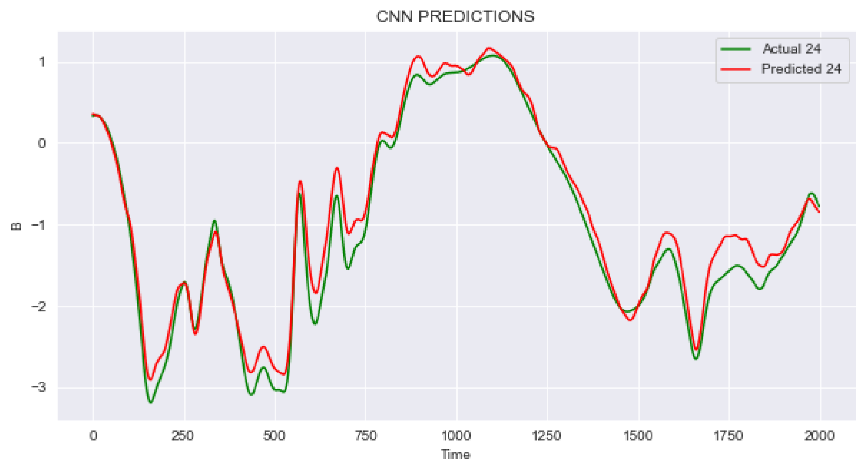
$$\frac{1}{c}\frac{dY_{j,k}}{dt} = \underbrace{-bY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k})}_{\text{Advection}} \underbrace{-Y_{j,k}}_{\text{Diffusion}} \underbrace{+\frac{h}{J}X_k}_{\text{Coupling}}$$

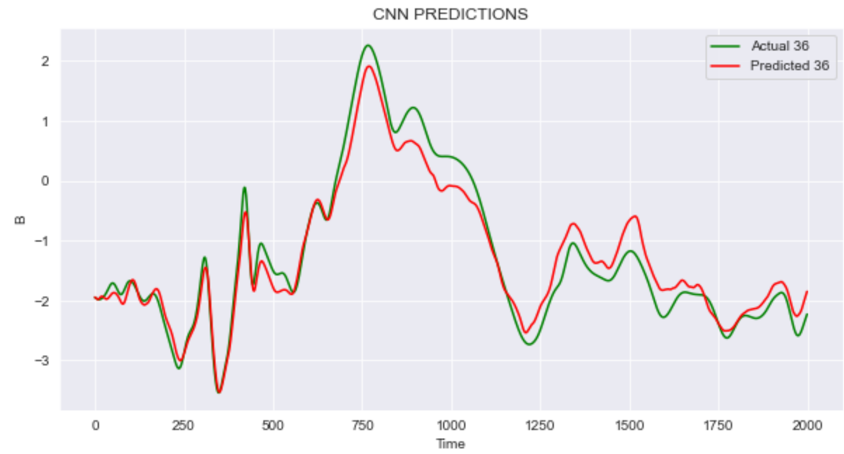
The parameterization is performed on the coupling term which is in turn dependent on the mean of all Y variables associated with a single X variable. In the previous implementation, I tried predicting one unknown variable X variable. Now my aim was to predict the coupling term for each X variable that dependent on another unknown variable Y. The dataset consisted of $K = 36$ variables of X and each X having 10 Y variables associated with it. A set of 10001 data points for the coupling parameter B was obtained using RK-4 method. The preprocessing of this dataset was done similar to the one discussed before for CNN model. The dataset was split into ratios of 80: 20 for training and testing. I used the same CNN model for prediction of the future data points for the coupling term B and the $\overline{Y}$ for each X variable. The initial conditions for X and Y were set.

Now, the training was performed with 50 epochs and the loss was set to RMSE. The average RMSE over 36 variables for the testing data was 0.002394119317517264 for B and 7.54575416208987e-05 for $\overline{Y}$. Given below are the plots for some variables representing the predicted data points for B. As previous results, the edge variables had higher RMSE values.

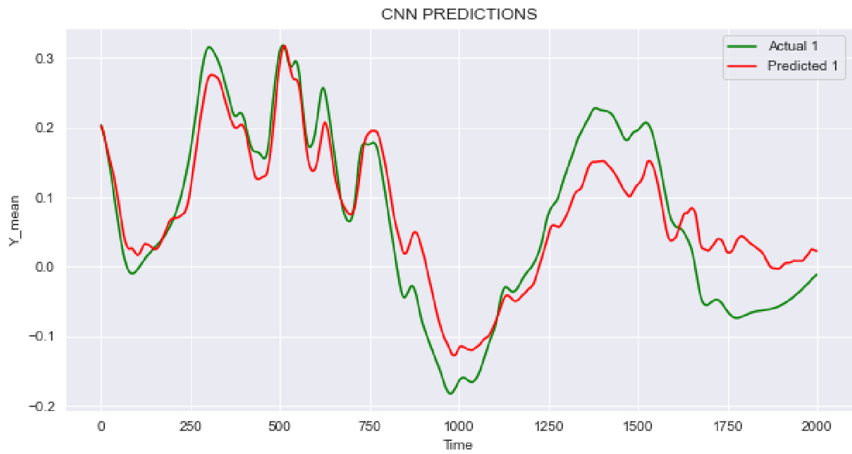


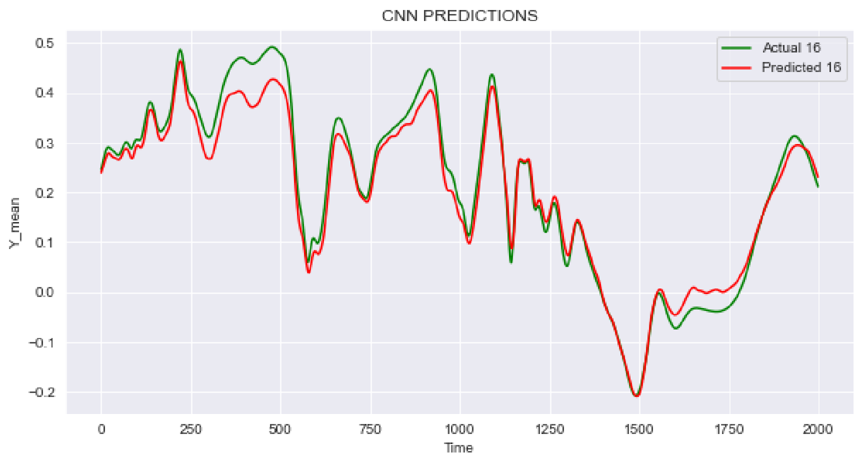


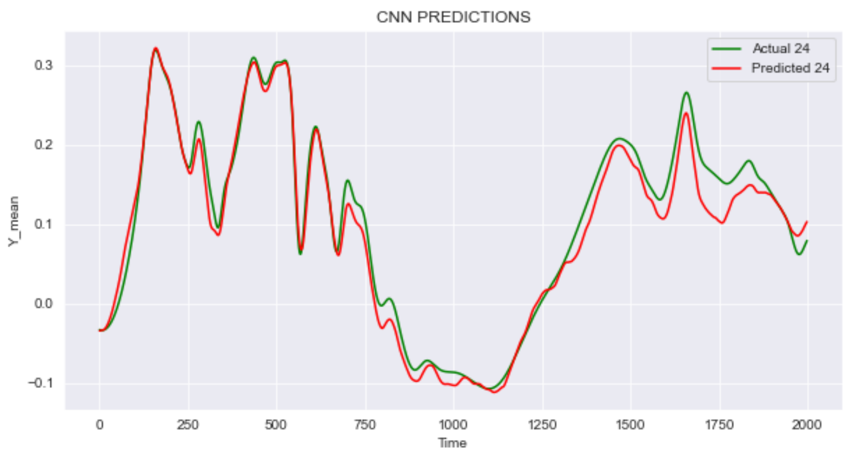


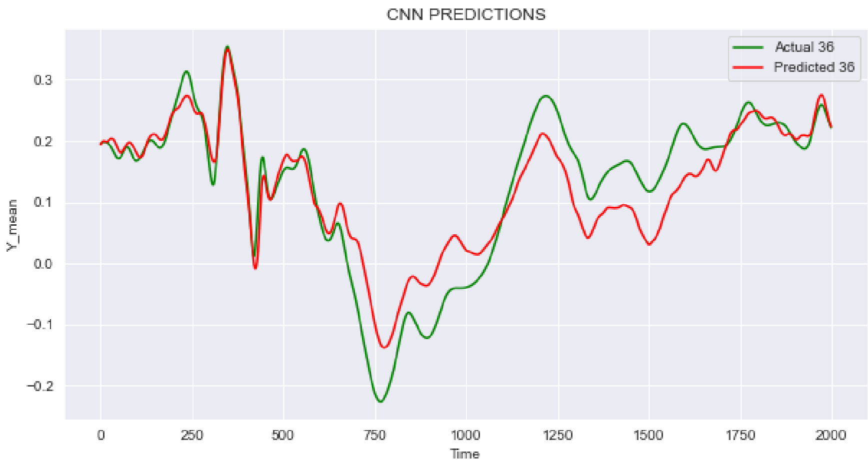


Next given below are the plots for prediction of the future data points for $\overline{Y}$ for some variables.









Conclusion

The Lorenz 96 is a simple yet powerful model that mimics the natural weather conditions. It has been extensively used in data assimilation and parameterization research. Recently, it has experienced a resurgence to test drive machine learning algorithms for parameter learning or sub-grid parameterizations. In this course, I tried to use ML/DL techniques for prediction of future data points for one unknown variable in Lorenz 96 model, and parameter learning for two unknown variables. I tried ARIMA model, which did not prove to be efficient as observed from the RMSE values and prediction failure in case of larger dataset. Then, I used a CNN model which showed significant decrease in the RMSE values between actual and predicted values. I used the same CNN model for one unknown variable prediction as well as parameterization or parameter learning in case of two unknown variables. For both cases, it was provided satisfactory results. It was a great experience to work under Dr. Zhaoxia Pu as I got an opportunity to experiment with parameterization in Lorenz 96 model and learn different techniques used for NWP.