OELP 2023

An efficient AI based Weather Forecasting System

Pitching presentation

Presented by : Priyanshu Gupta, 112001033
Om Tiwari, 132001027

Mentor : Dr. Satyajit Das
Dr. Sreenath Vijayakumar

OELP 2023

Problem Statement

To develop real time weather forecasting system to predict weather conditions for the next 14 days.

Our Solution

Choose the best IMF values for a particular feature to feed along with other features in the LSTM to get prediction of that particular feature for next 14 days.

Divide each feature into its components which we name as IMFs.

Actual Dataset

	DDATETIME	AVGT	AVGP	MAXP	MINP	MINT	MAXT	MINU	WNDSPD
0	2015-01-01	165	10166	10195	10128	127	204	24	21
1	2015-01-02	148	10172	10191	10148	117	203	28	19
2	2015-01-03	155	10147	10176	10118	118	216	23	17

01 Actual dataset that we are dealing with has 20 features which we reduced to 8 as some features were highly dependent (which we can see using kmo test).

02 KMO is a test conducted to examine the strength of the partial correlation (how the factors explain each other) between the variables.

03 We got Shenzhen (China) dataset to start on our project which had data of nearly past 5000 days from which we filtered out the same 8 features.

OELP 2023

KMO test on 7 features

AVGT	AVGP	MAXP	MINP	MINT	MAXT	MINU
0.757	0.759	0.834	0.849	0.814	0.817	0.679

01 A correlation test between variables can be realized by the KMO test. The closer the KMO statistic is to 1, the stronger the correlation between variables is, and the weaker the partial correlation is.

02 The KMO test values of all variables are greater than 0.6, indicating that there is a high correlation between all-weather variables

03 Multivariable prediction has a reliable basis.

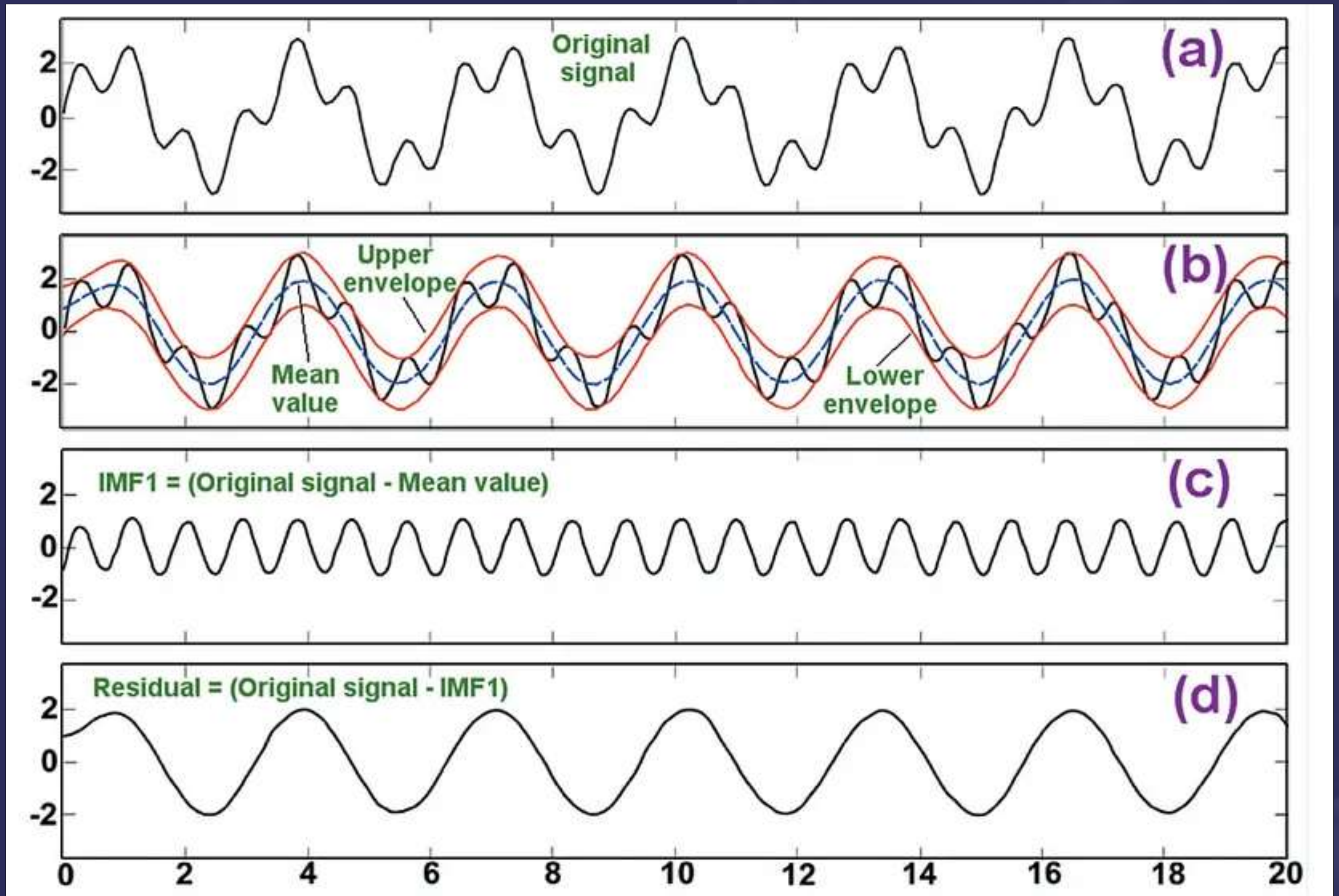
OELP 2023

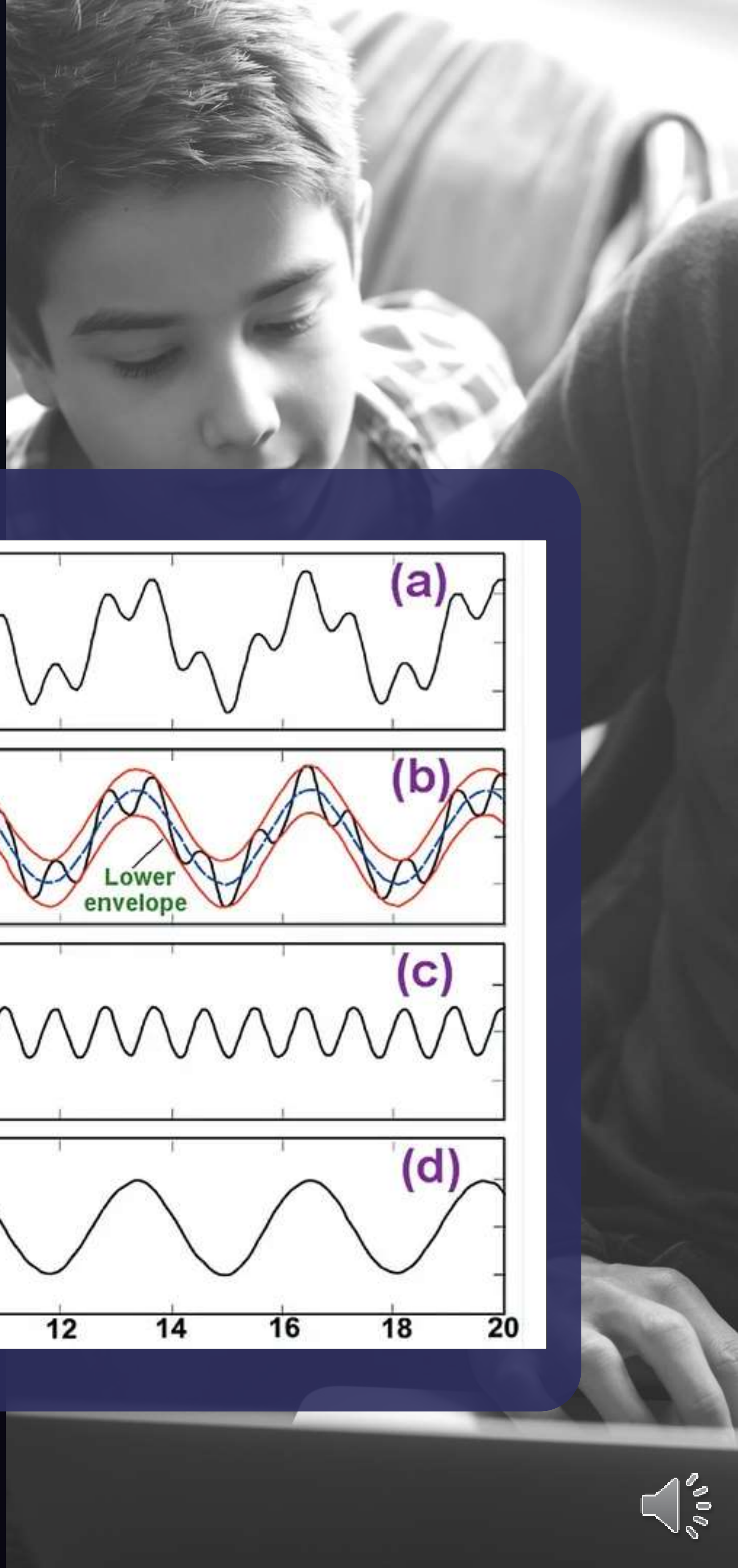
EMD is a method of breaking down a signal(any time series data into IMFs) without leaving the time domain.

Two conditions for IMF –

- The difference number of maxima and minima at most by 1.
- The Mean of the wave of IMF is zero.

Pre-Processing of Data





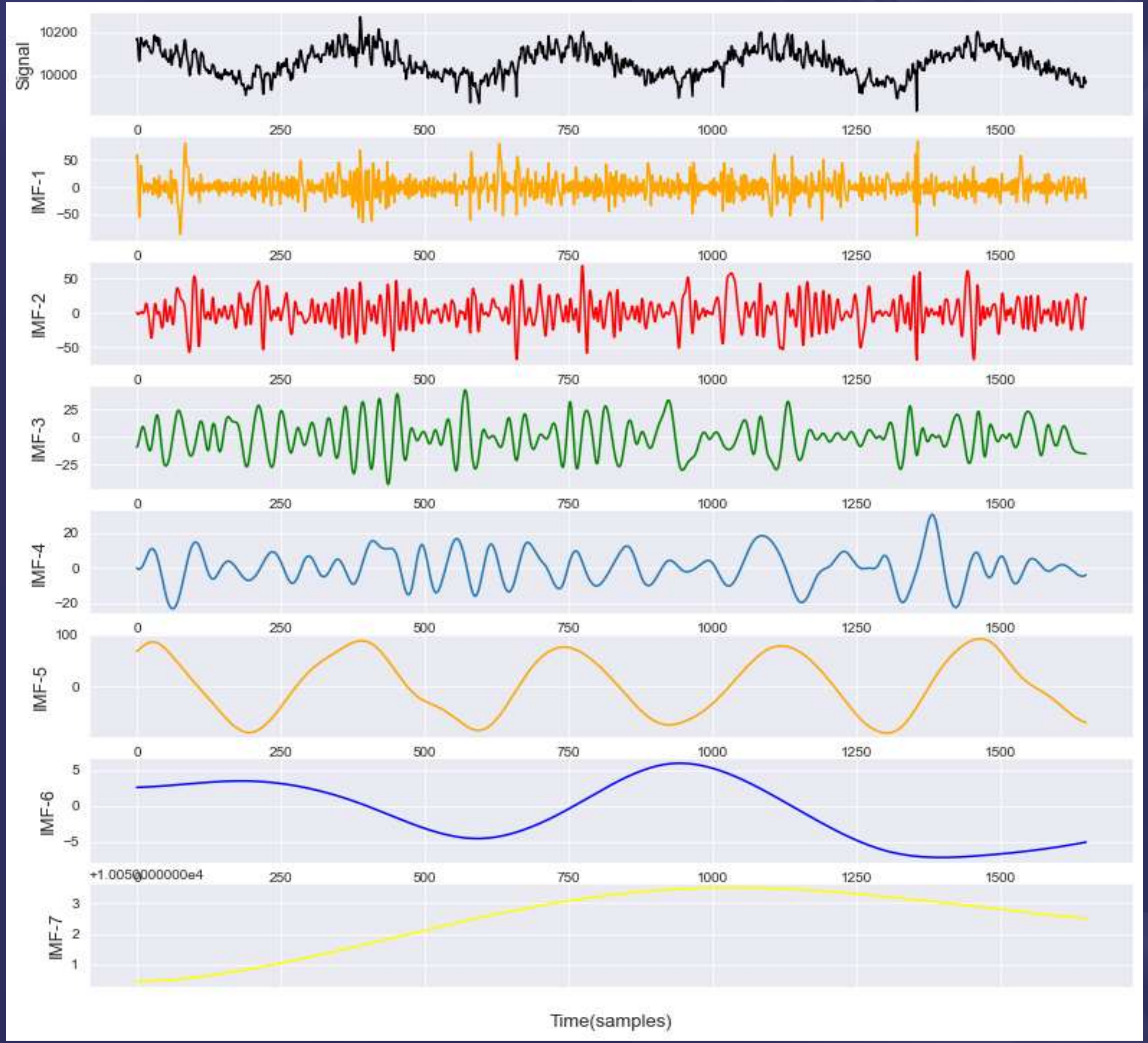
OELP 2023

IMF distribution of Average Pressure for (Shenzhen Data) for first 1649 days.

Two conditions for IMF –

- The difference number of maxima and minima at most by 1.
- The Mean of the wave of IMF is zero.

Empirical Mode Decomposition



AVGP high correlation with IMF-5 (0.864)

Pearson's correlation

1 Describe the correlation of average pressure with its IMF values

OELP 2023

LSTM Model

	AVGT	AVGP	MAXP	MINP	MINT	MAXT	MINU	avgp_imf5
0	165	10166	10195	10128	127	204	24	68.135283
1	148	10172	10191	10148	117	203	28	69.192951
2	155	10147	10176	10118	118	216	23	70.242387
3	172	10094	10124	10071	134	232	48	71.280462
4	199	10065	10085	10050	181	229	67	72.304737

Final Training Dataset

01 For all the features that we are interested in we are feeding all the features along with the best IMF values (values more than 0.5) of the specific feature.

02 We are achieving loss of 0.04 by single variate prediction which is 10 times lesser if we perform training for multivariate output.

03 Doing validation on 0.1 split of training dataset, giving us loss of 0.0022 while training with 0.0018 on validation set and mse of $2 * 10^{-6}$ after inverse scaling of data for last 50 days.

OELP 2023

Training the Model

```
Layer (type)                 Output Shape              Param #
=================================================================
lstm_19 (LSTM)               (None, 7, 50)             11800

lstm_20 (LSTM)               (None, 7, 50)             20200

lstm_21 (LSTM)               (None, 50)                20200

dense_7 (Dense)              (None, 8)                 408
=================================================================
Total params: 52,608
Trainable params: 52,608
Non-trainable params: 0
```

Lstm Model

01

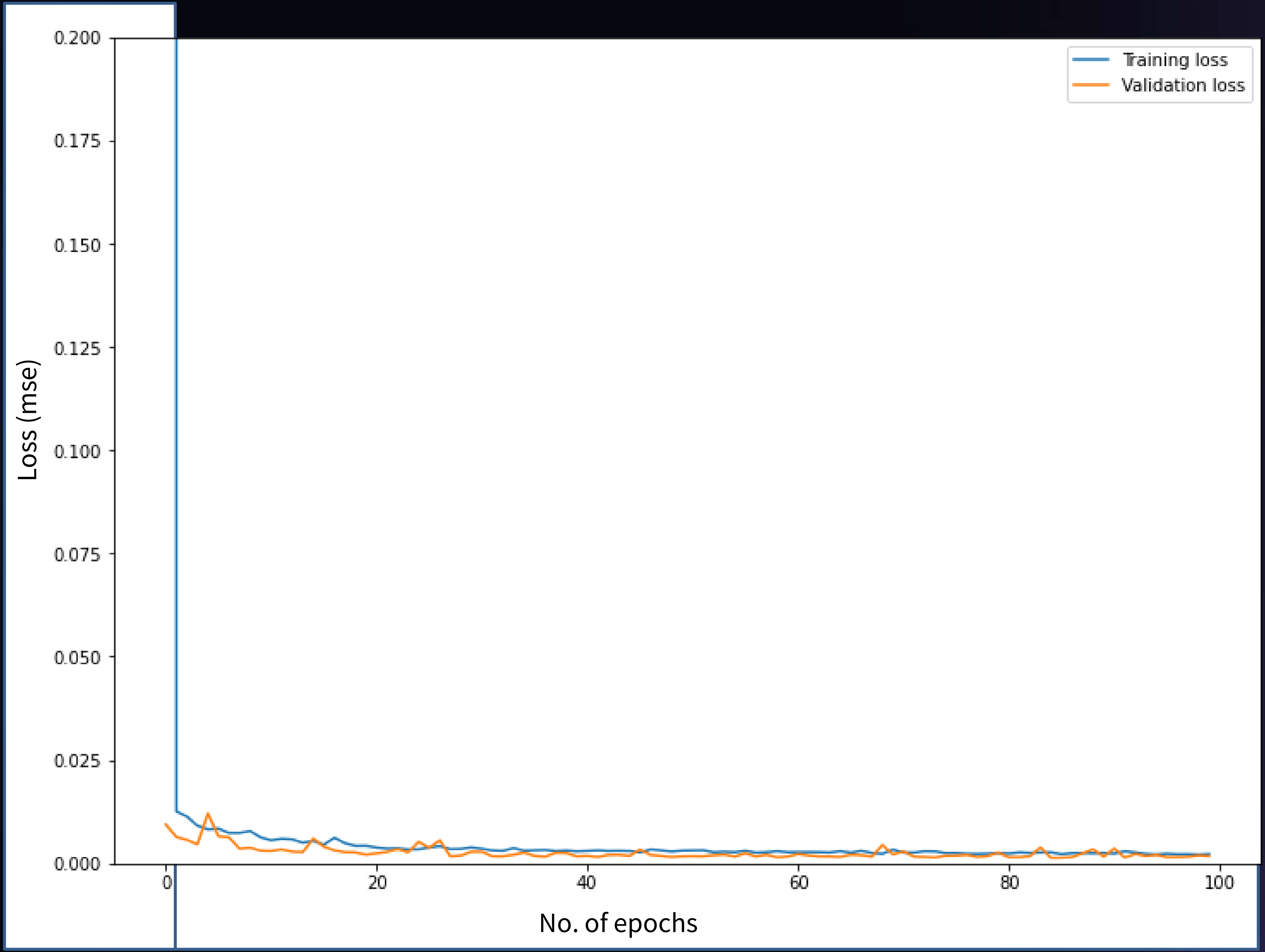
n_future = 1 # Number of days we want to look into the future based on the past days.
n_past = 7 # Number of past days we want to use to predict the future.

02

- trainX shape == (1642, 7, 8).
- trainY shape == (1642, 1).

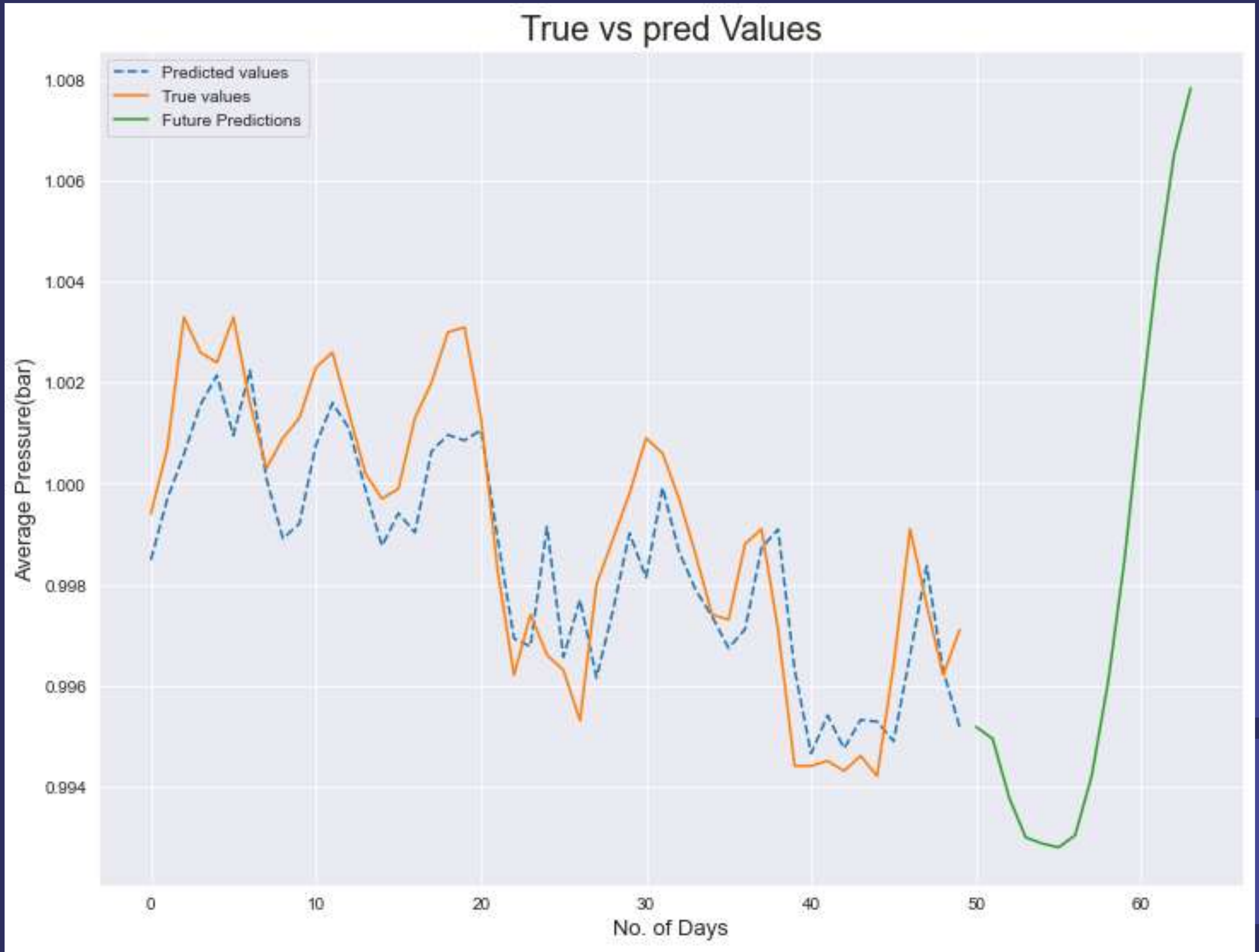
03

Feeding all the features along with the best IMF values (corresponding to that feature) .



Training and Validation Loss

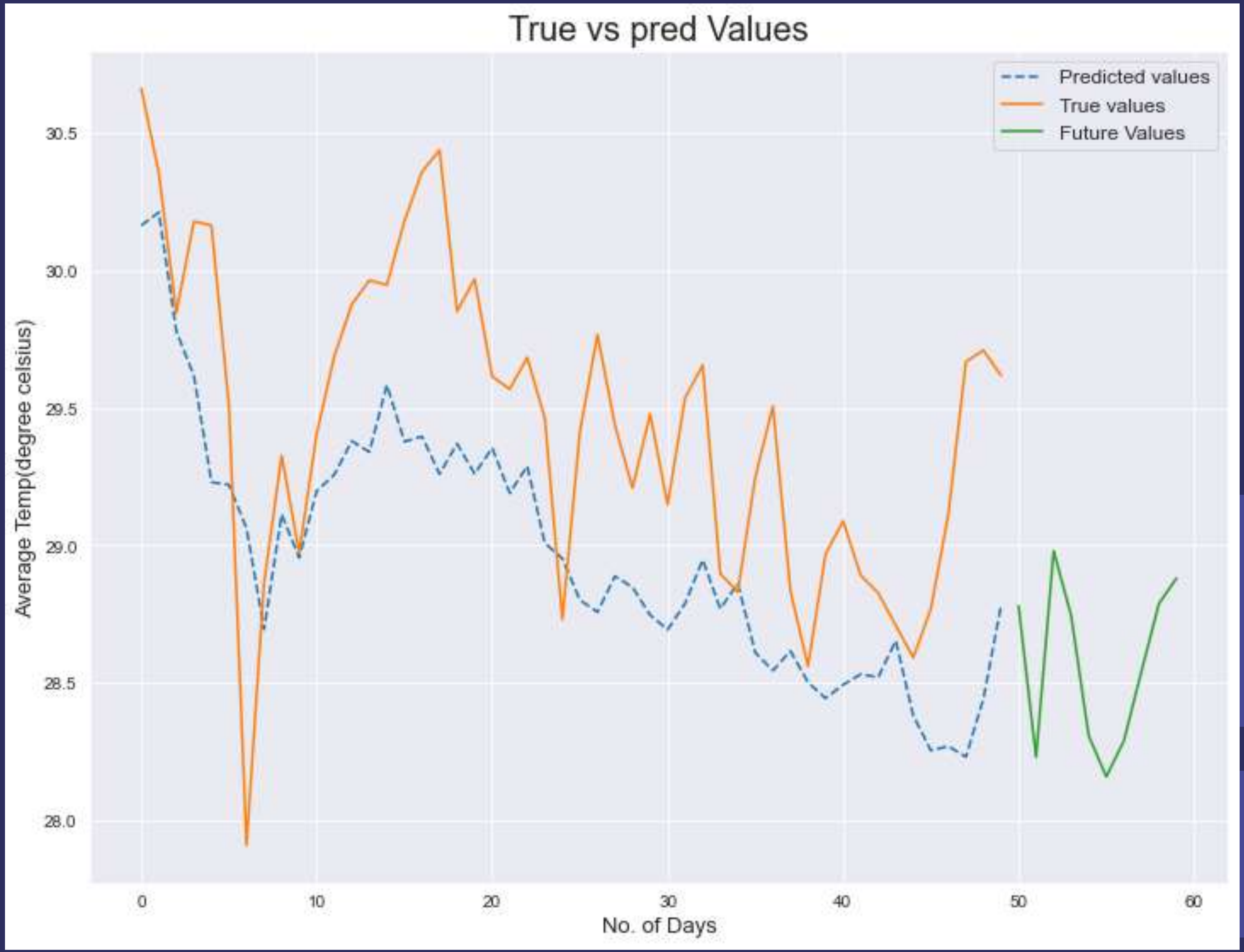
OELP 2023



Prediction and Forecasting for AVG Pressure (shenzhen data)

1. Training loss of 0.0022 while validation loss of 0.0018 and mse of $2 * 10^{-6}$ on original data (inverse scaled)
2. Trained on 1642 days.

OELP 2023



Prediction and Forecasting for AVG Temp (Trivandrum data)

1. Training loss of 0.063 while validation loss of 0.062 and mse of 0.39 on original data (inverse scaled)
2. Trained on 303 days only

THANK YOU

References

- https://www.mdpi.com/2073-4433/13/8/1208
- https://www.researchgate.net/publication/282408199_A_hybrid_method_combining_Teager_Kaiser_energy_operator_empirical_mode_decomposition_and_minimum_entropy_deconvolution_for_monitoring_gears_damages