ECAL Seq2Seq Learning

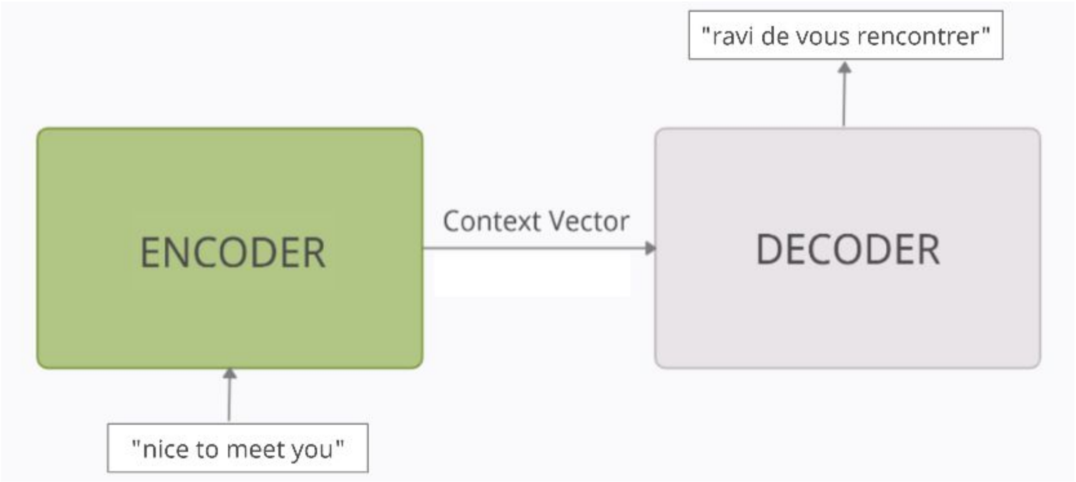
Seq2Seq Model

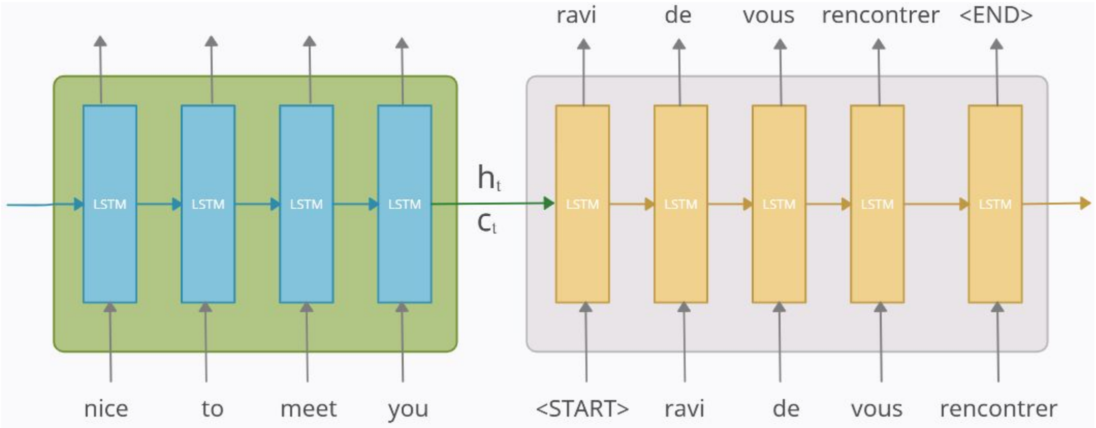
Input: (Enighlish) “nice to meet you”
Output: (French) “ravi de vous rencontrer”

Encoder: Processing each token in the input-sequence & encoding all the information about the input-seq into a fixed length vector.

Context vector: Encapsulating the whole meaning of the input-seq that can help the decoder make accurate predictions.

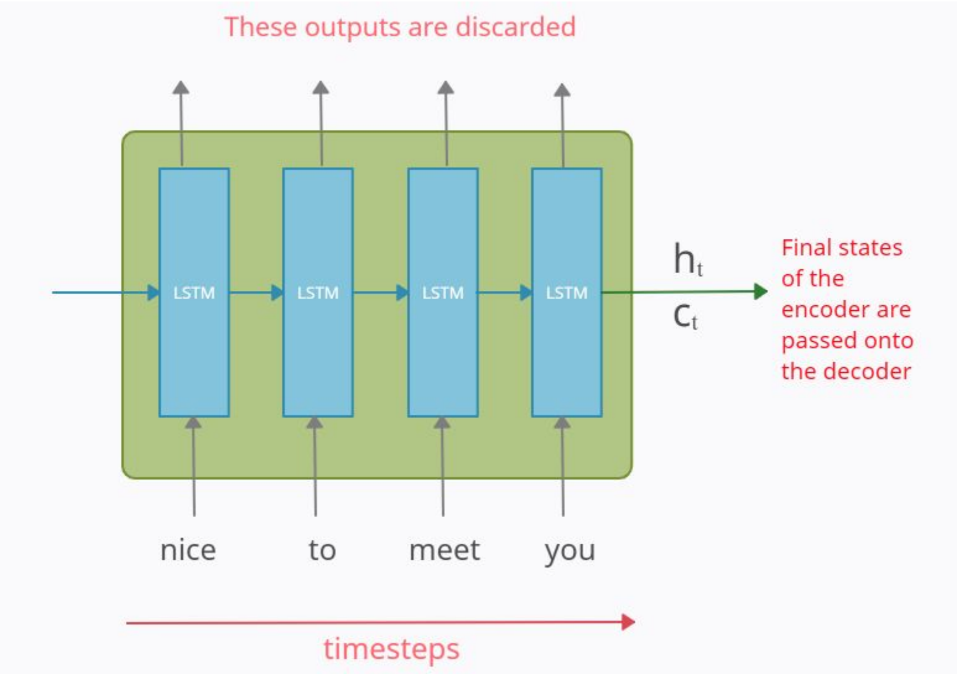
Decoder: Reading the context vector and tries to predict the target-seq token by token.

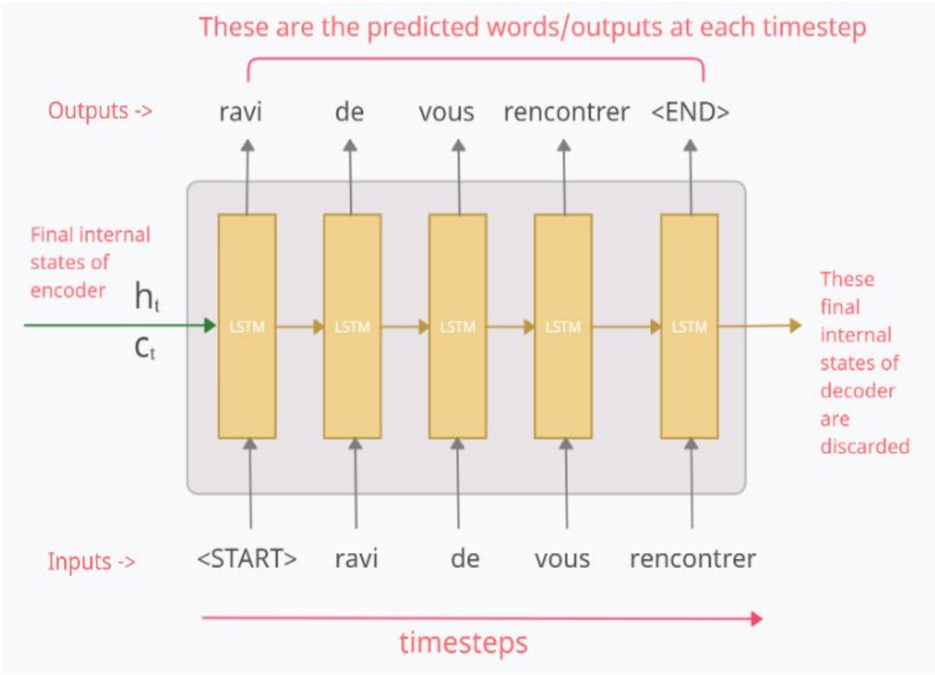




Ref:
https://medium.com/analytics-vidhya/encoder-decoder-seq2seq-models-clearly-explained-c34186fbf49b

Seq2Seq Model





Encoder **Decoder**

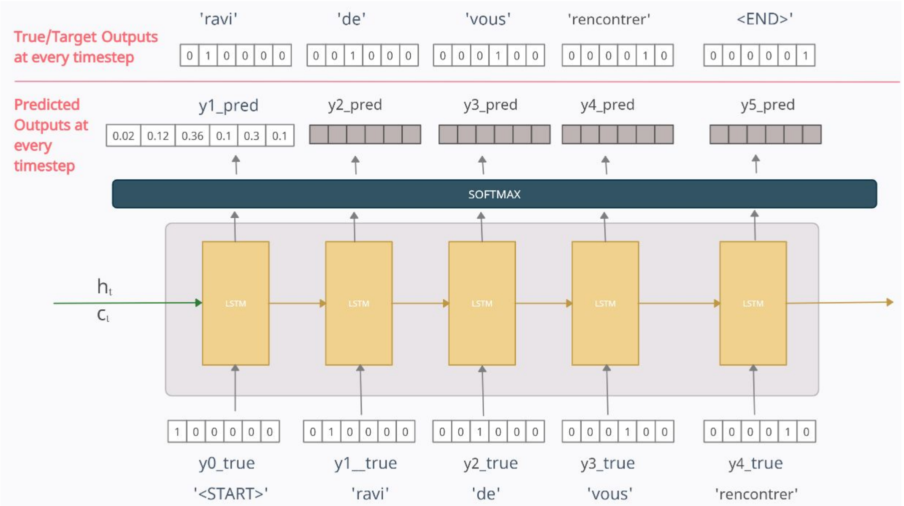
Ref: https://medium.com/analytics-vidhya/encoder-decoder-seq2seq-models-clearly-explained-c34186fbf49b

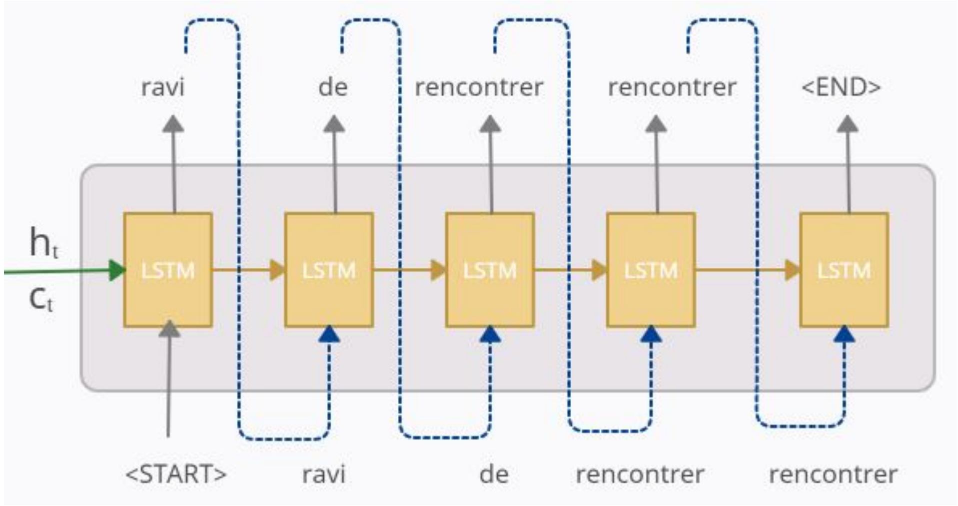
Seq2Seq Training & Test

The Decoder in Training Phase:

1) **Teacher Forcing:** feeding the **true token** (and not the predicted output/token) from the previous time-step as input to the current time-step.

2) **Without teacher forcing:** using its own predictions as the next input

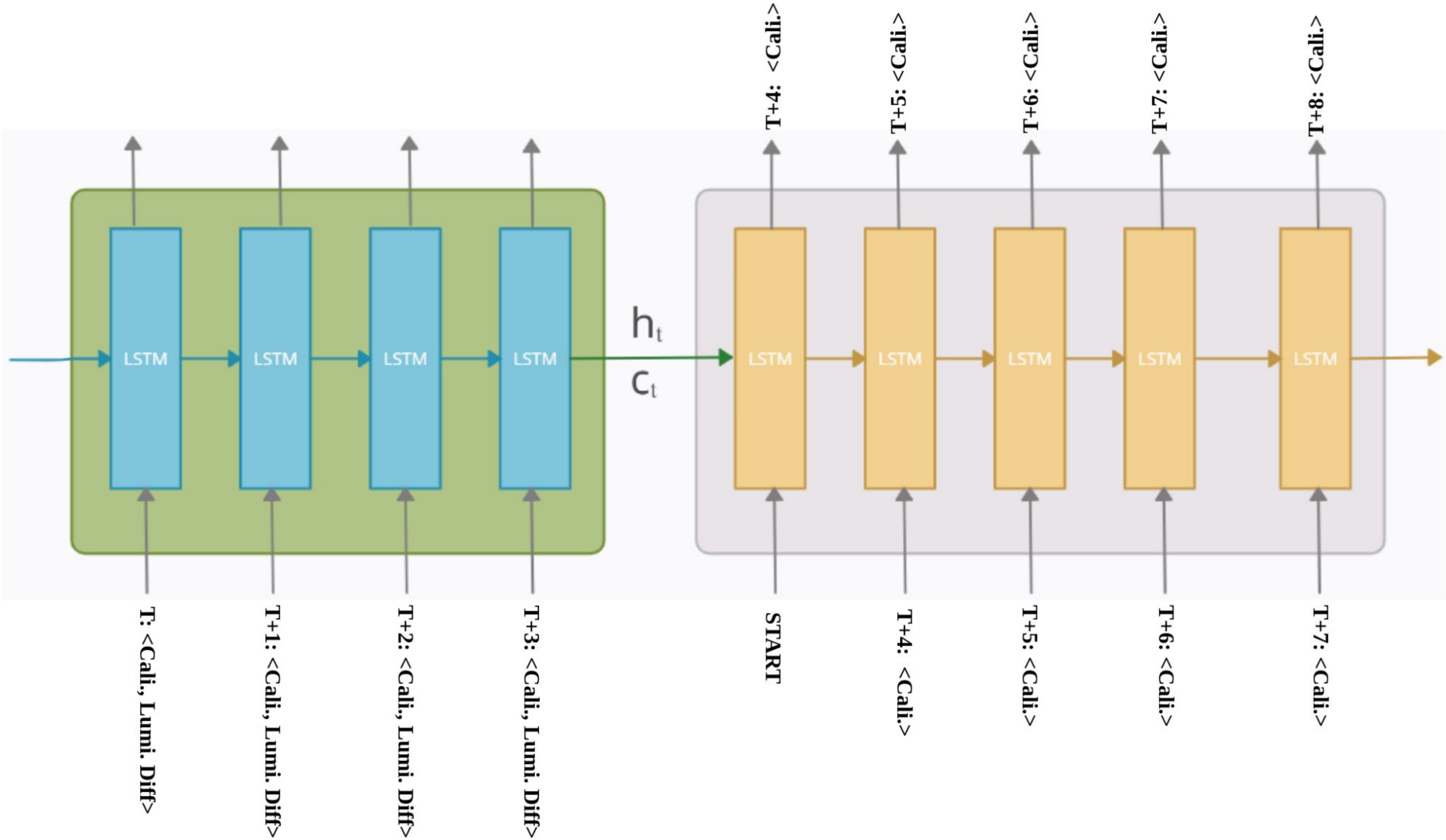




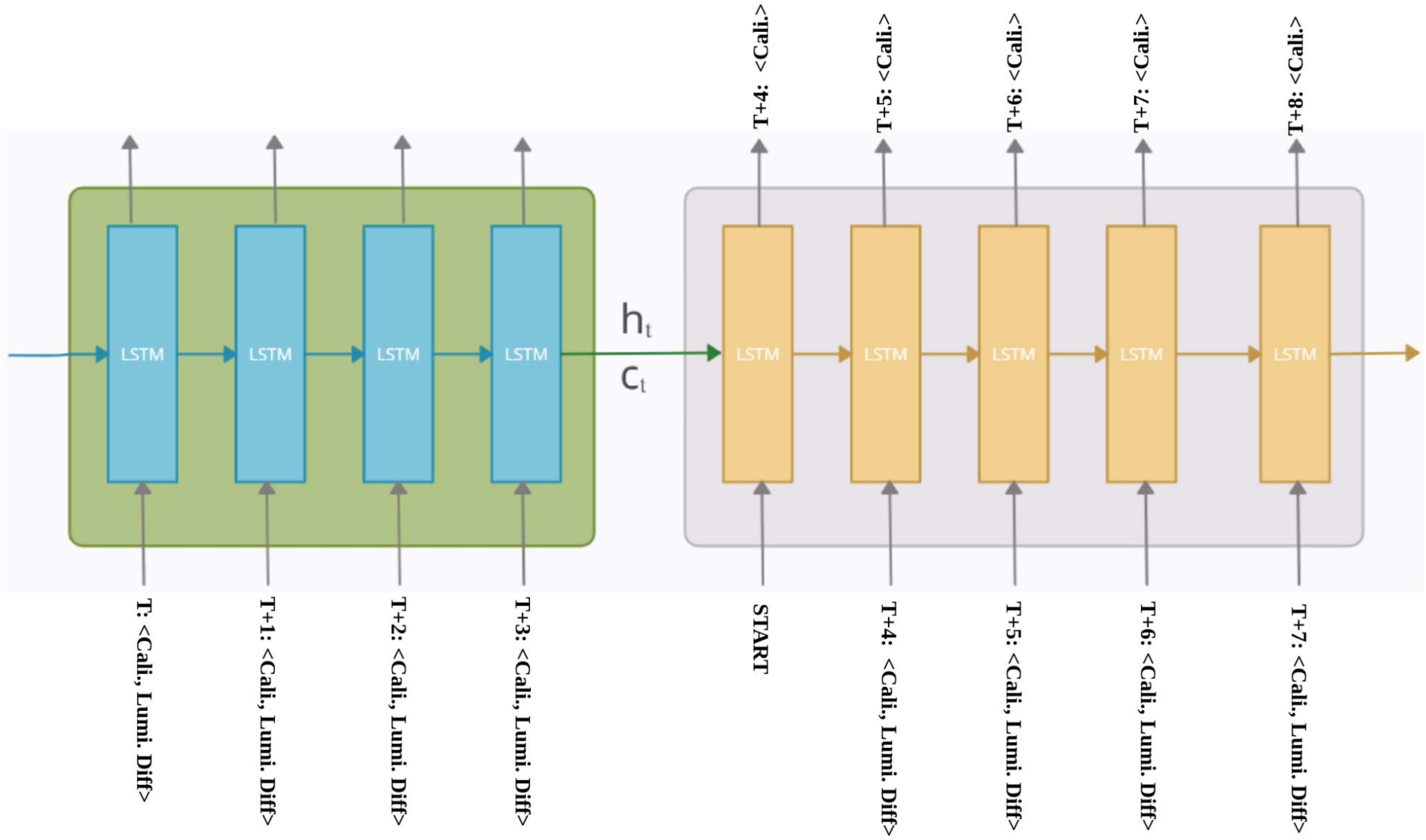
Teacher Forcing

Without Teacher Forcing

Our Seq2Seq Model Type-1



Our Seq2Seq Model Type-2



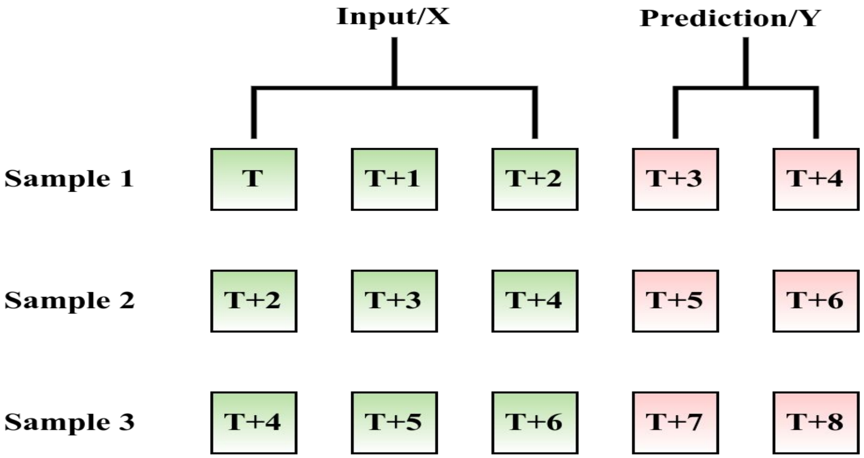
Training/Test Data Format

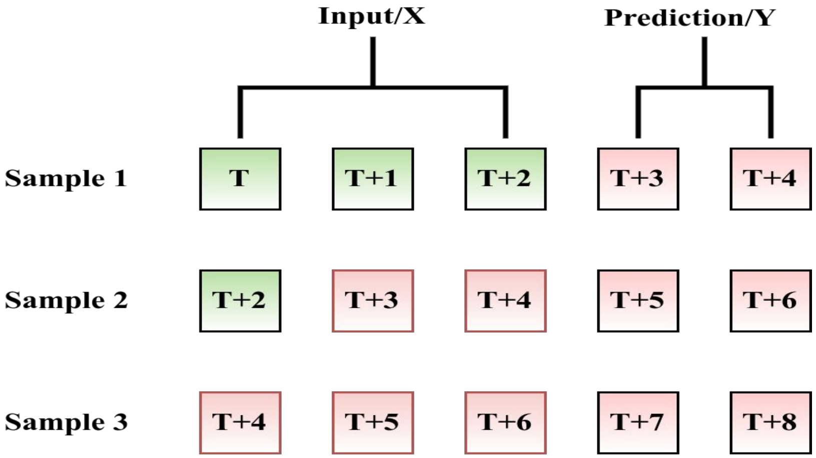
Case1 (left):

1) We can always observe 3 consecutive actual values and then make predict on the next two values;
2) When we predict "T+3 & T+4", we use the actual "T, T+1, T+2";
3) When we want to predict "T+5 & T+6", we wait until we obtained the actual "T+3 & T+4".

Case 2 (right):

1) The only observed information we have is "T, T+1, T+2";
2) In order to make much further prediction, we need to "re-use" our prediction as "fake observation".



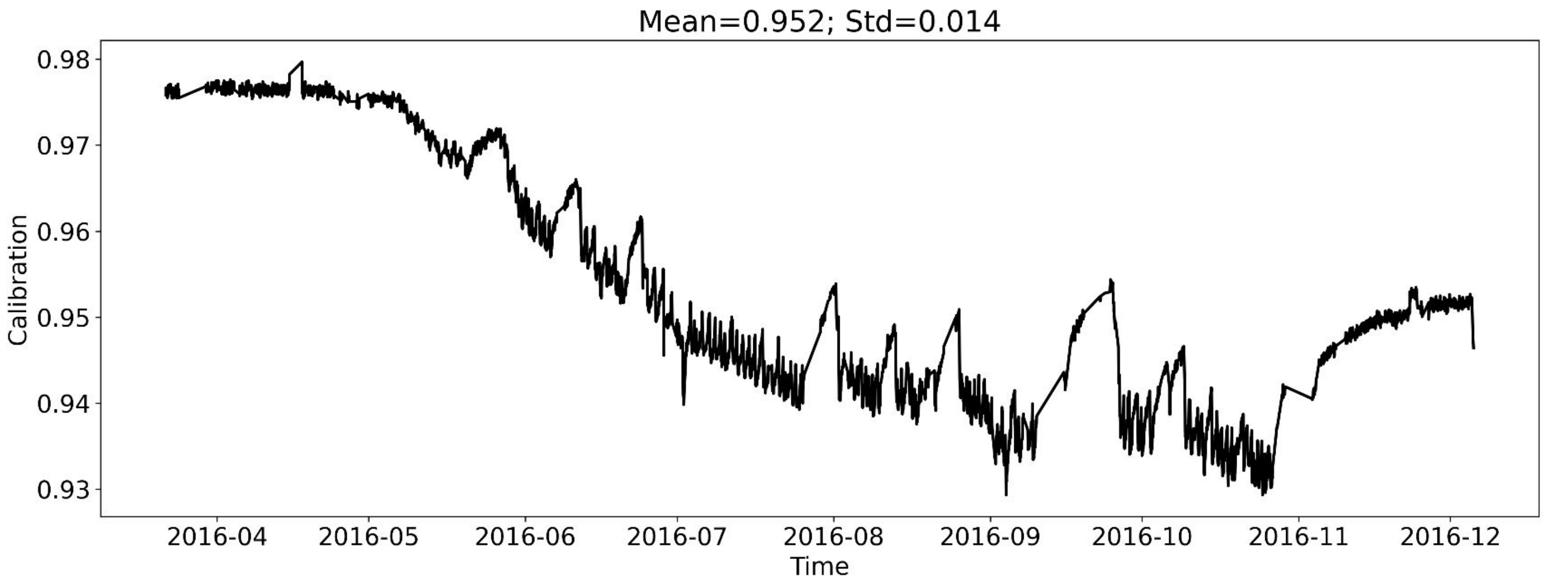


Experimental Setting Up

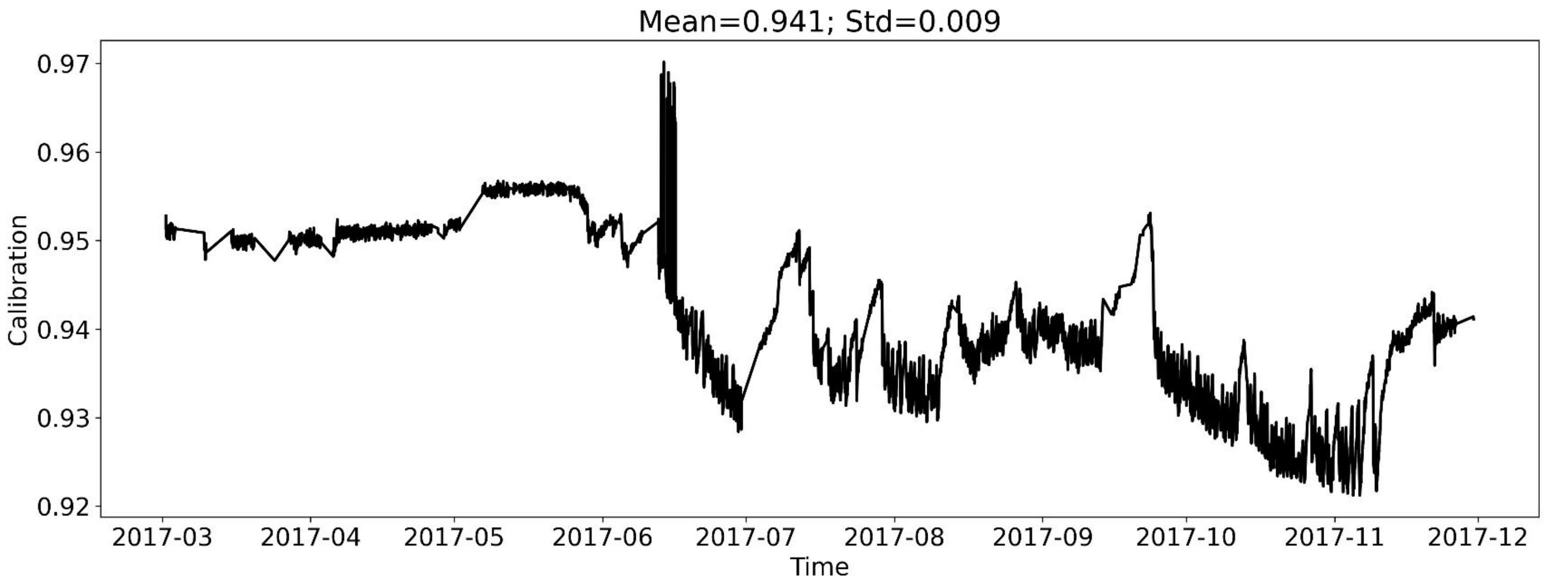
All results in the following slides use the same setting up:

1) We use Seq2Seq Model Type-2 (see slide 6 for details);
2) We use Case 1 (see slide 7 for details);
3) We train our model on 2016 data of 54000 crystal; and we test the trained model on 2017 data, 2o18 data of 54000 crystal.

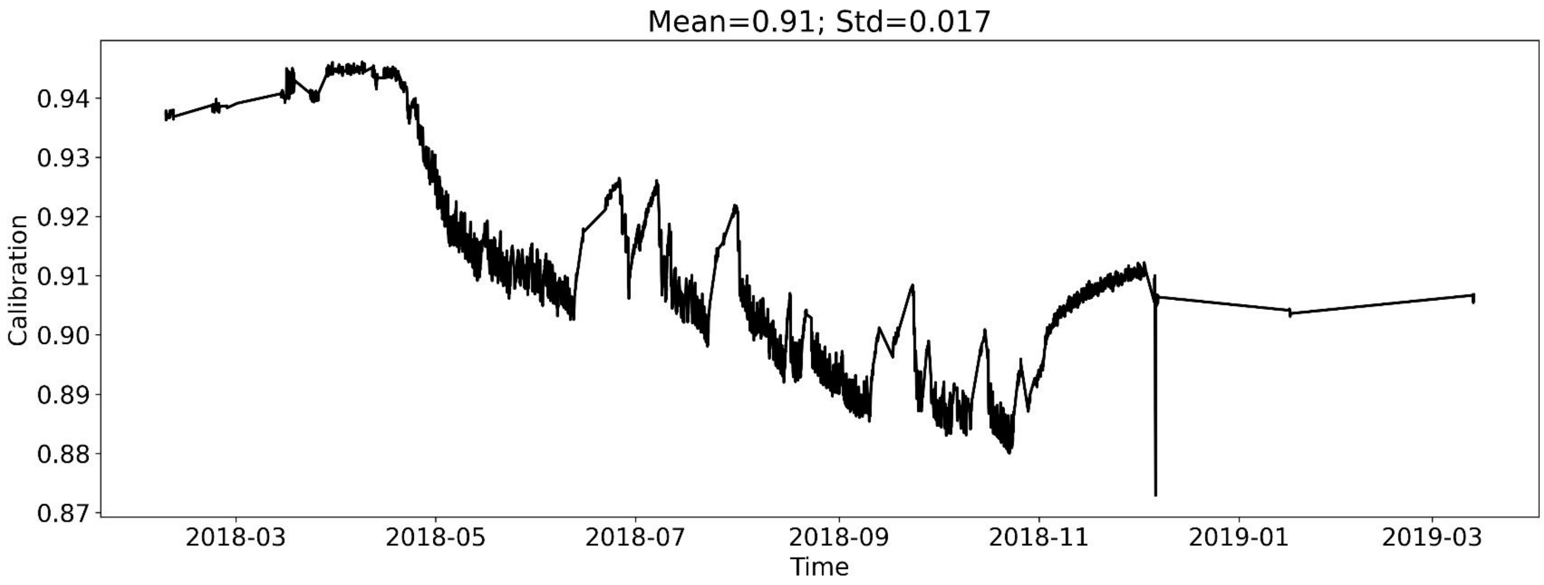
Original Calibration-2016



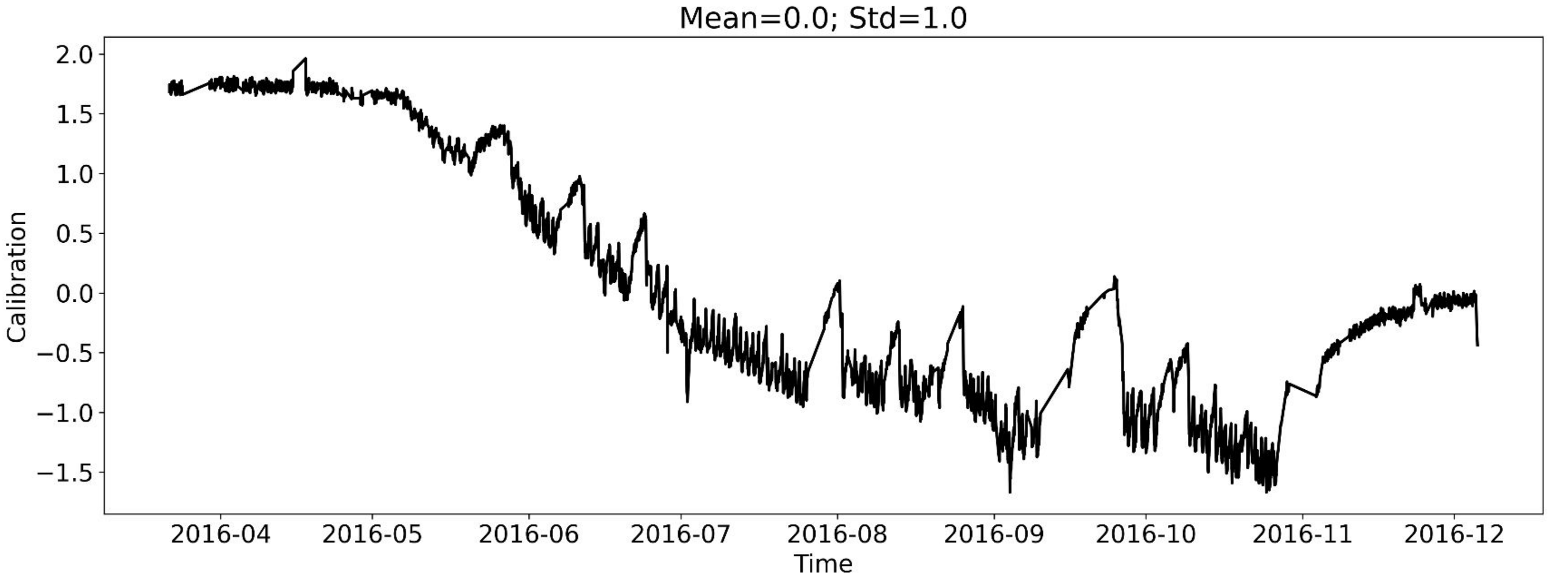
Original Calibration-2017



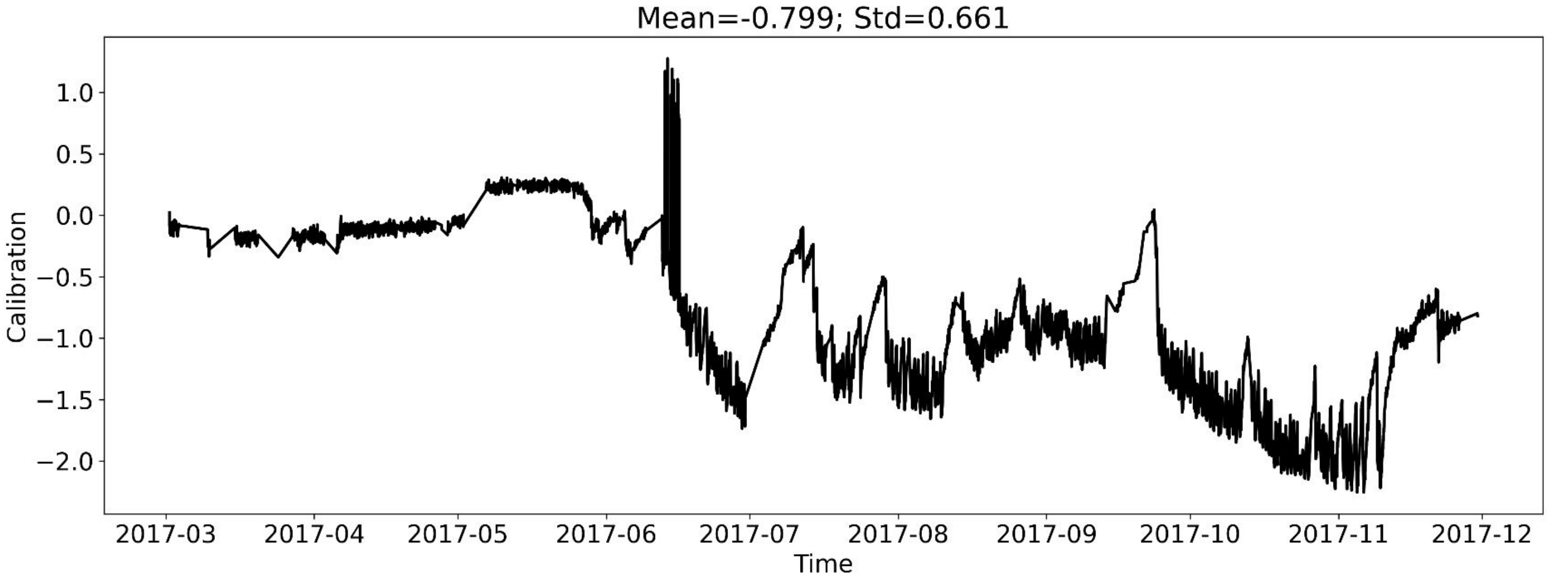
Original Calibration-2018



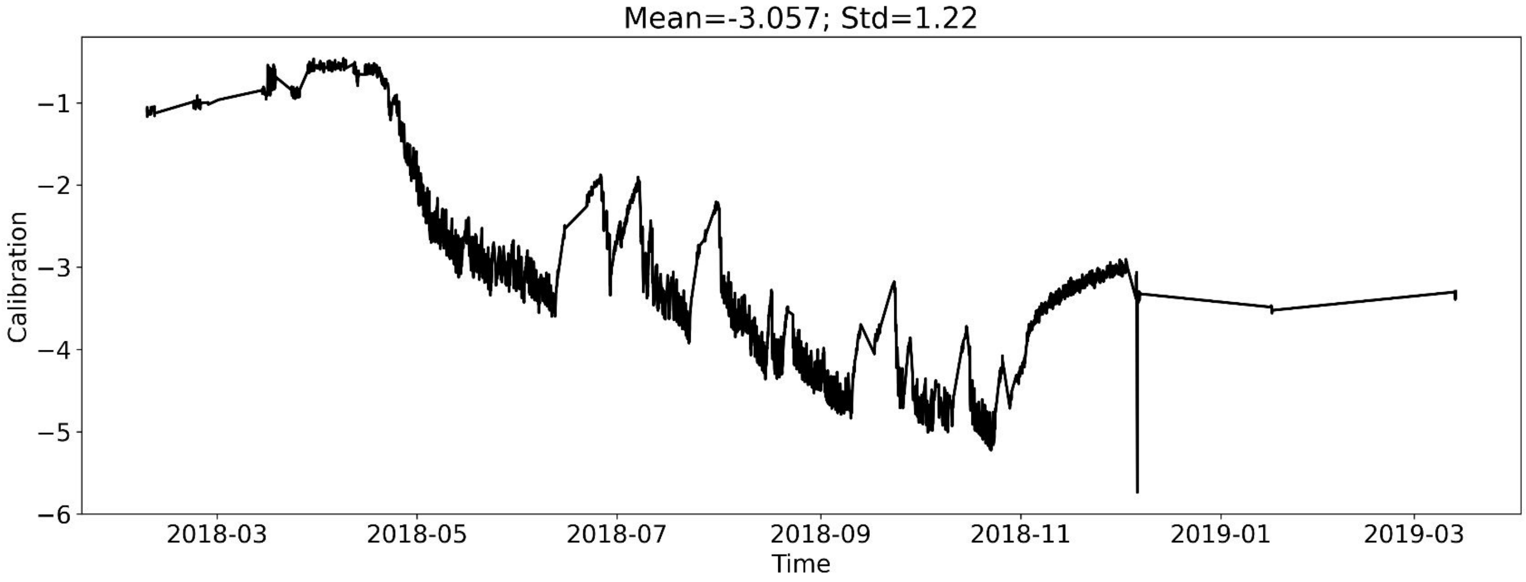
Normalized Calibration-2016

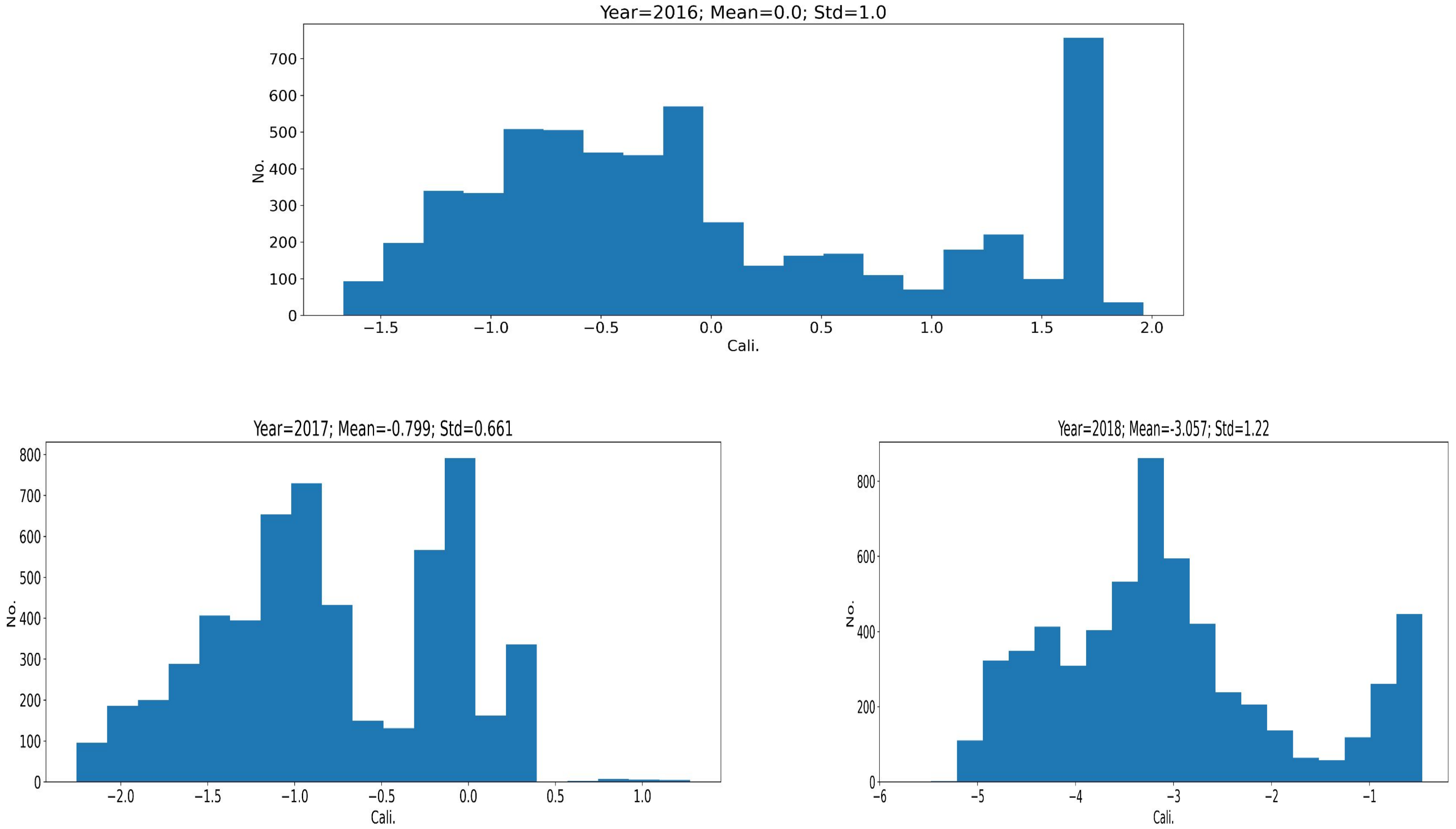


Normalized Calibration-2017

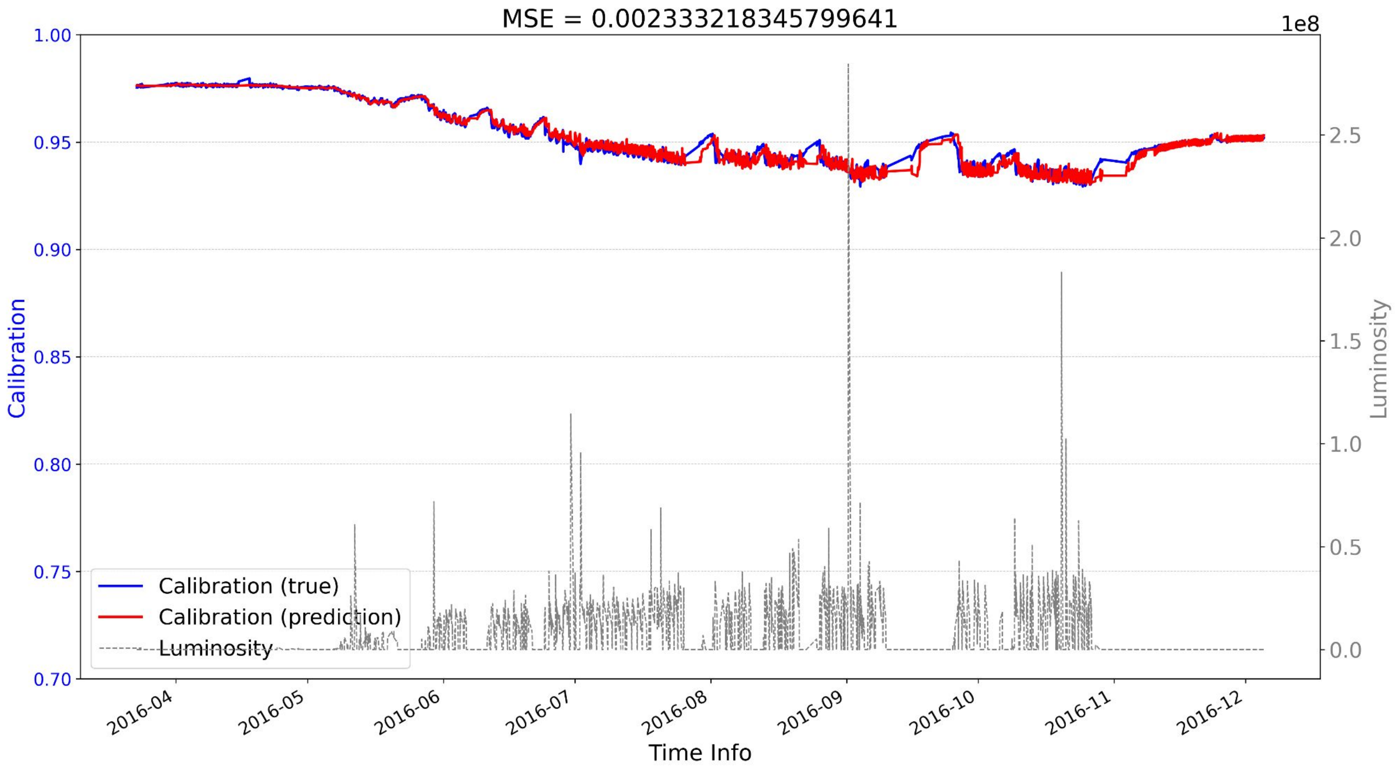


Normalized Calibration-2018

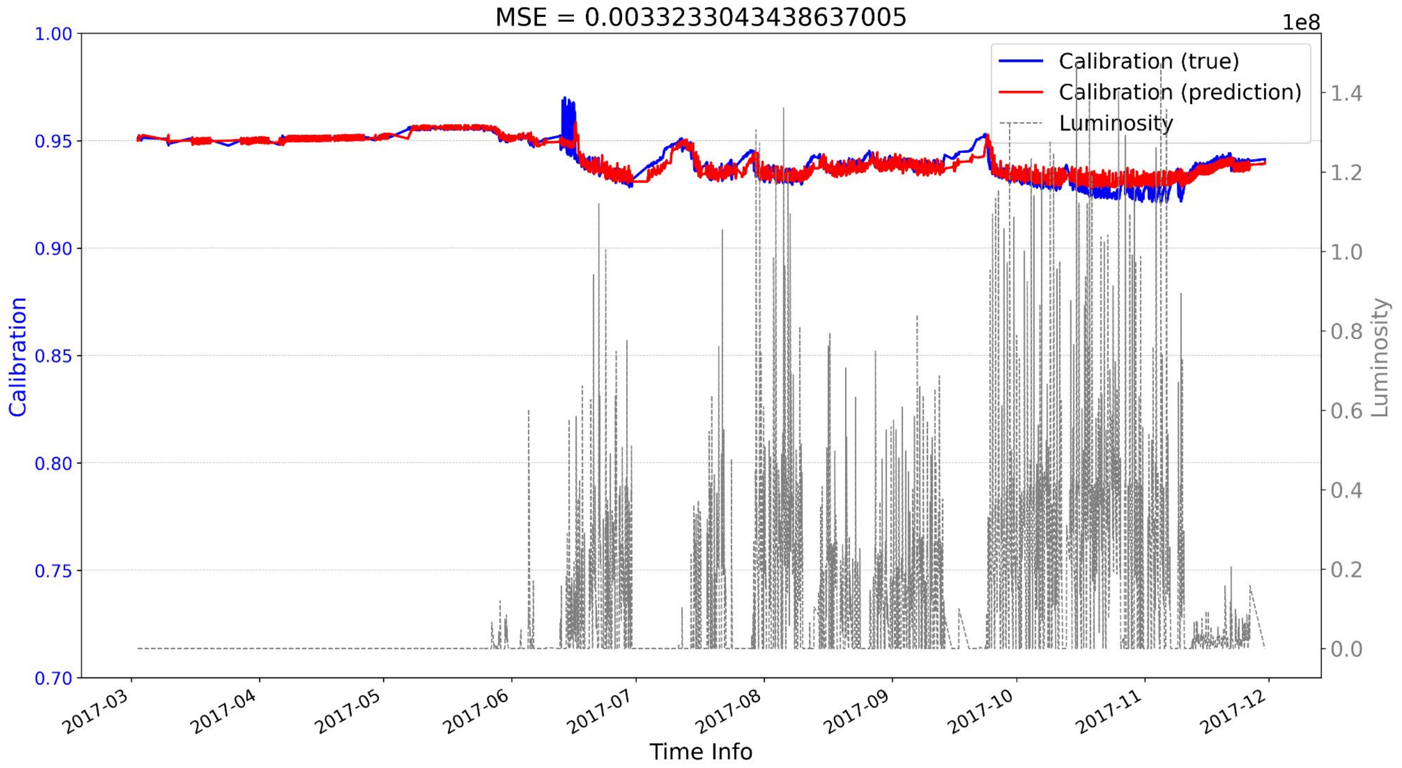



Year=2016; Mean=0.0; Std=1.0
No.
Cali.
Year=2017; Mean=-0.799; Std=0.661
No.
Cali.
Year=2018; Mean=-3.057; Std=1.22
No.
Cali.

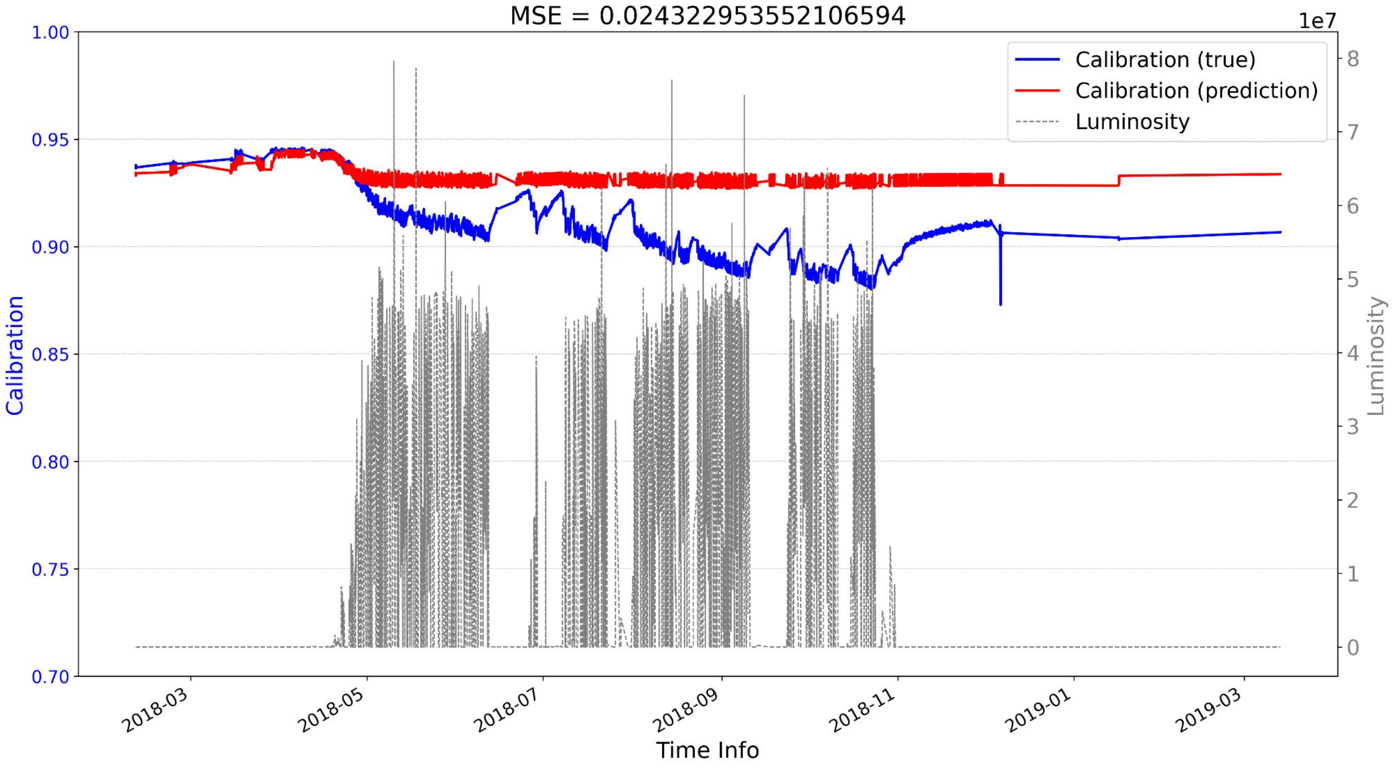
Results—Training 0n 2016; Test on 2017 & 2018



Results—Training 0n 2016; Test on 2017 & 2018



Results—Training 0n 2016; Test on 2017 & 2018

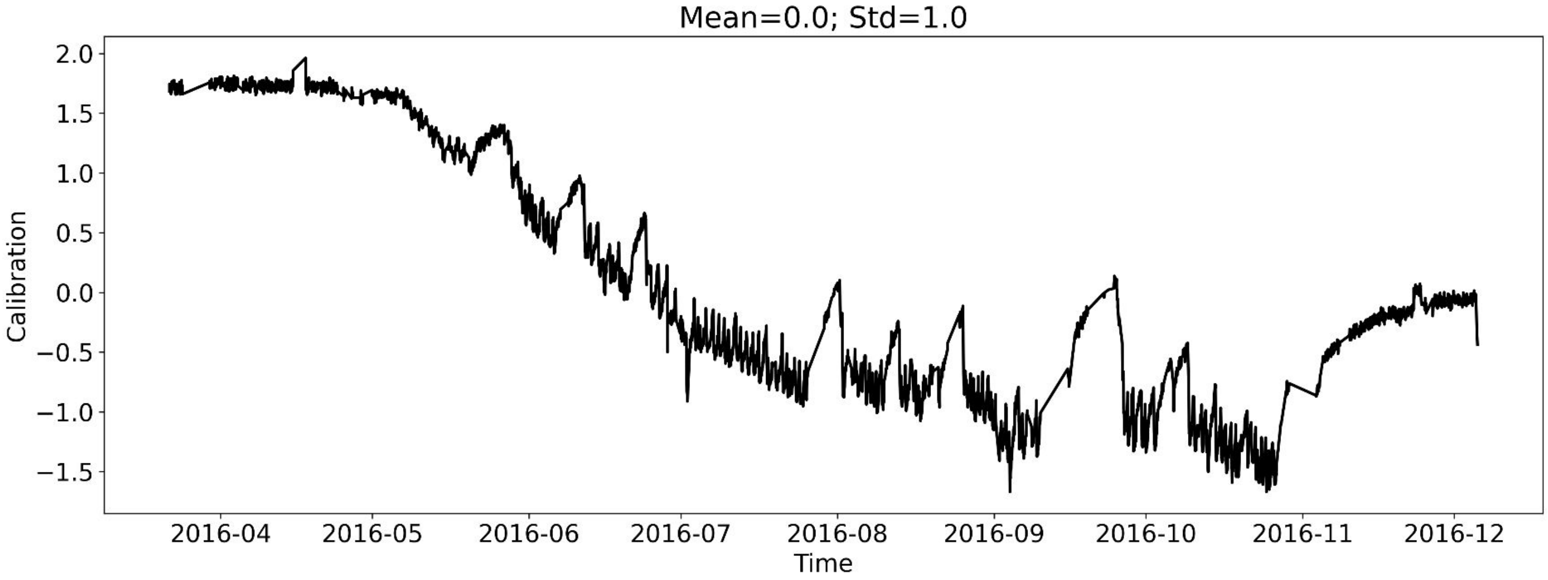


Results Analysis

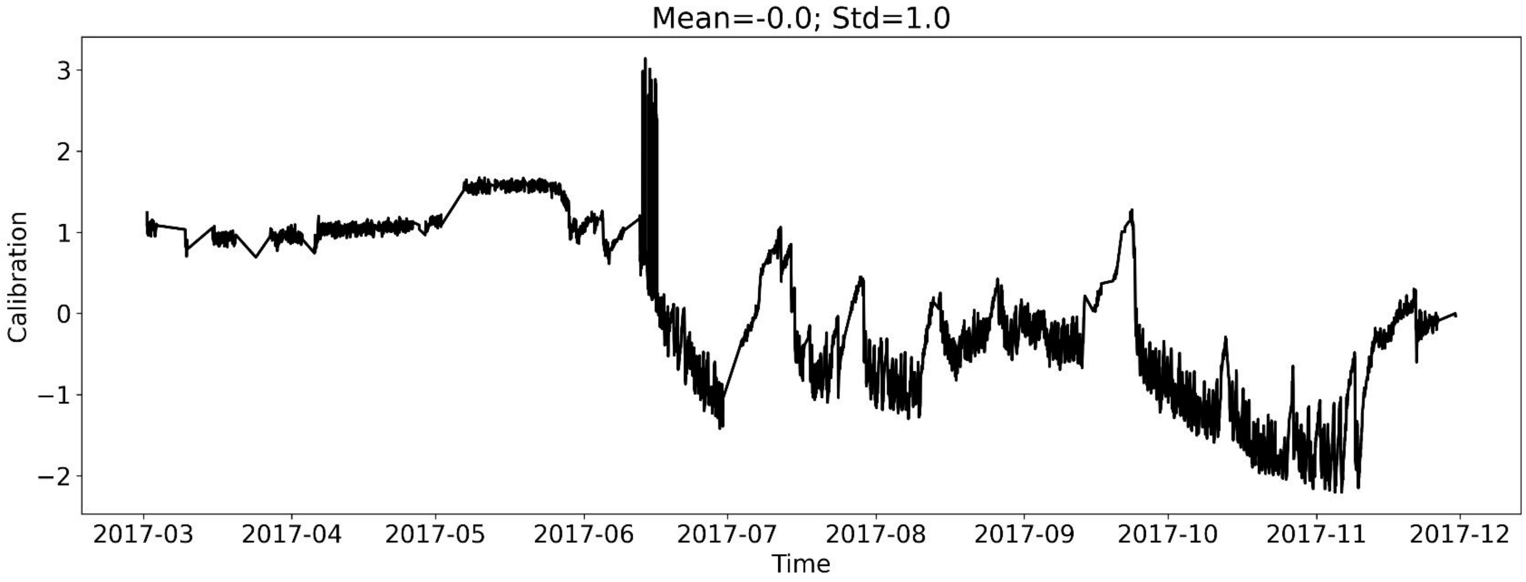
One potential reason that causes the prediction performance degradation:

1) Data distribution shift

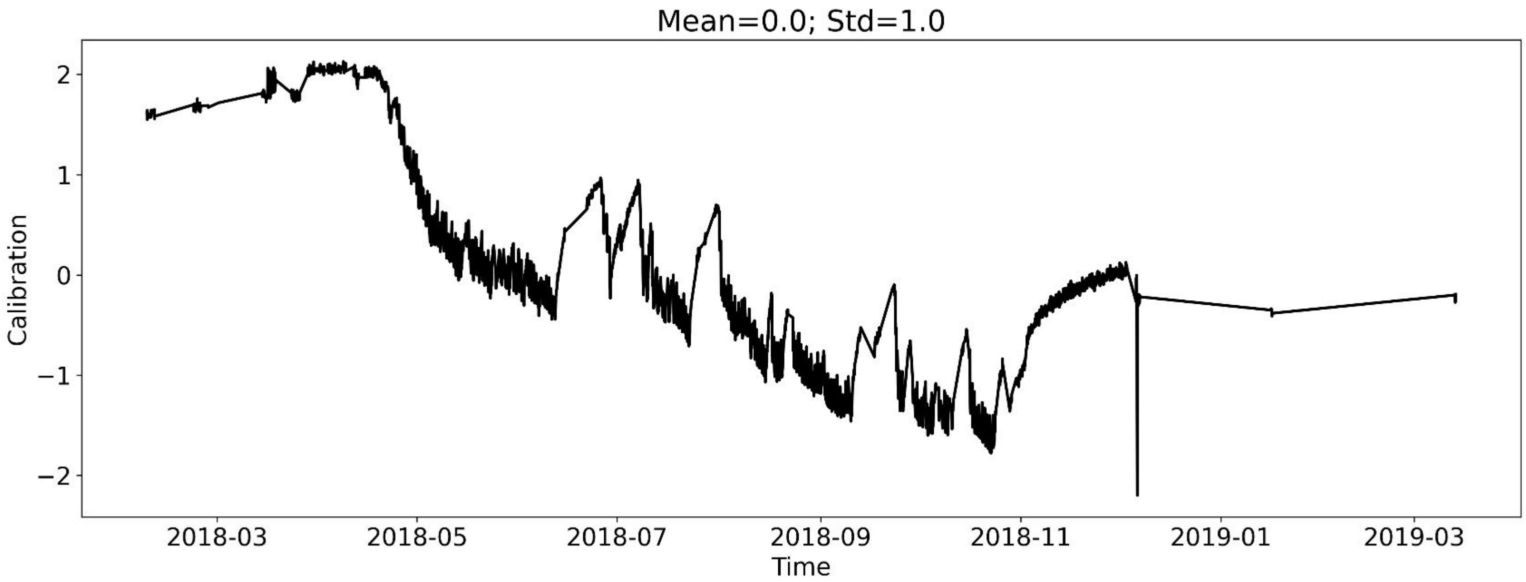
If we normalize the data separately-2016

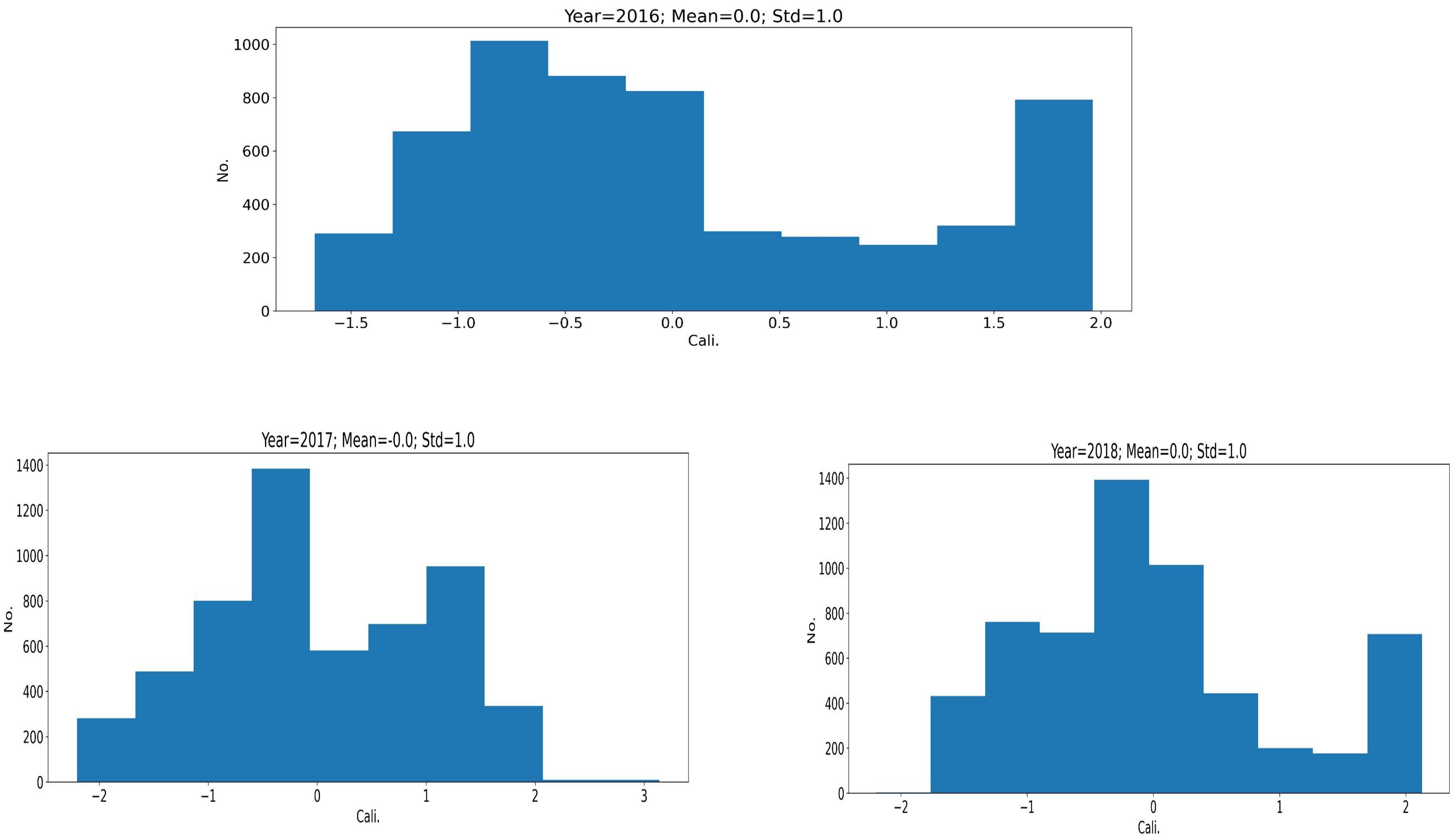


If we normalize the data separately-2017

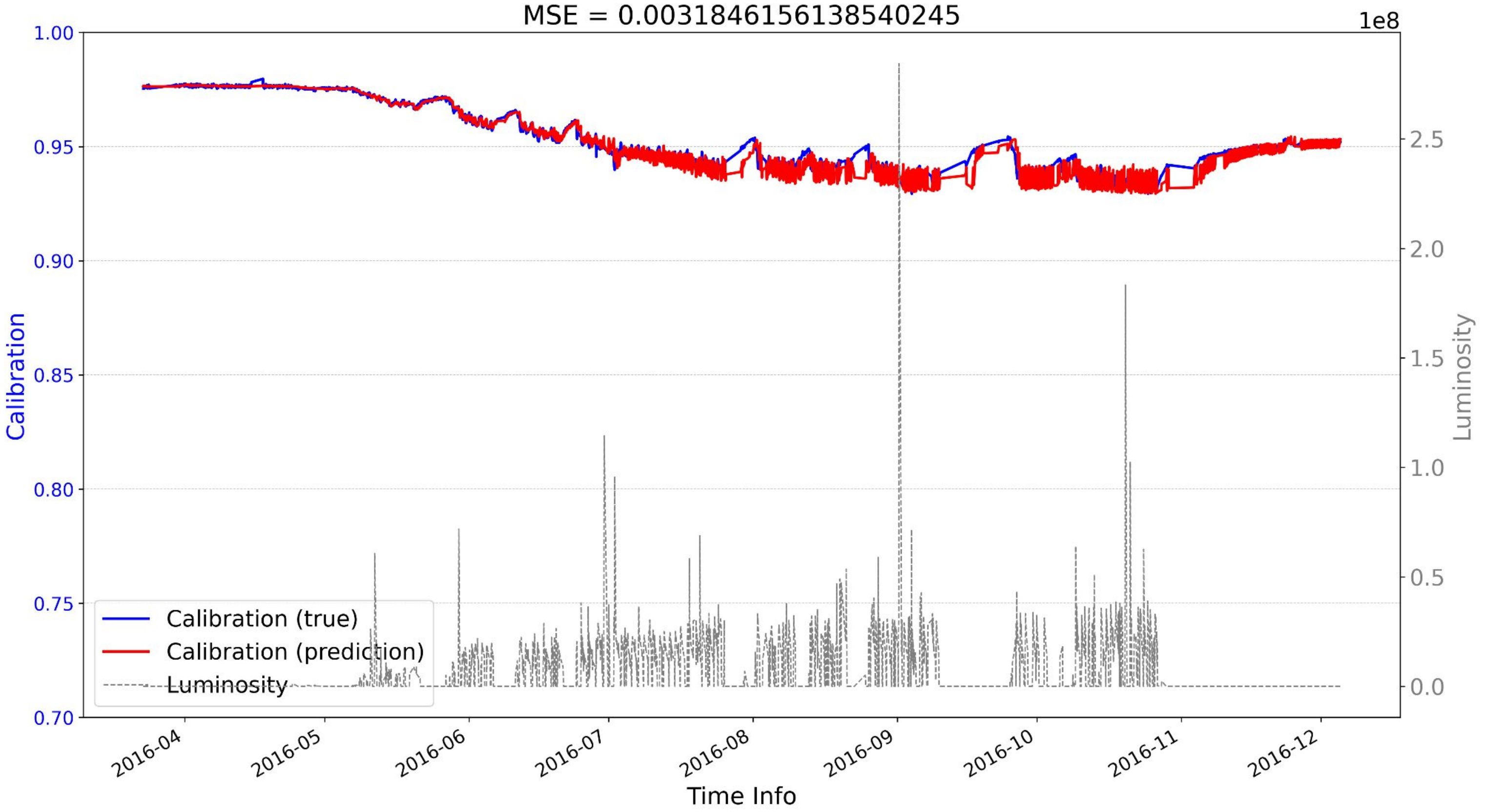


If we normalize the data separately-2018

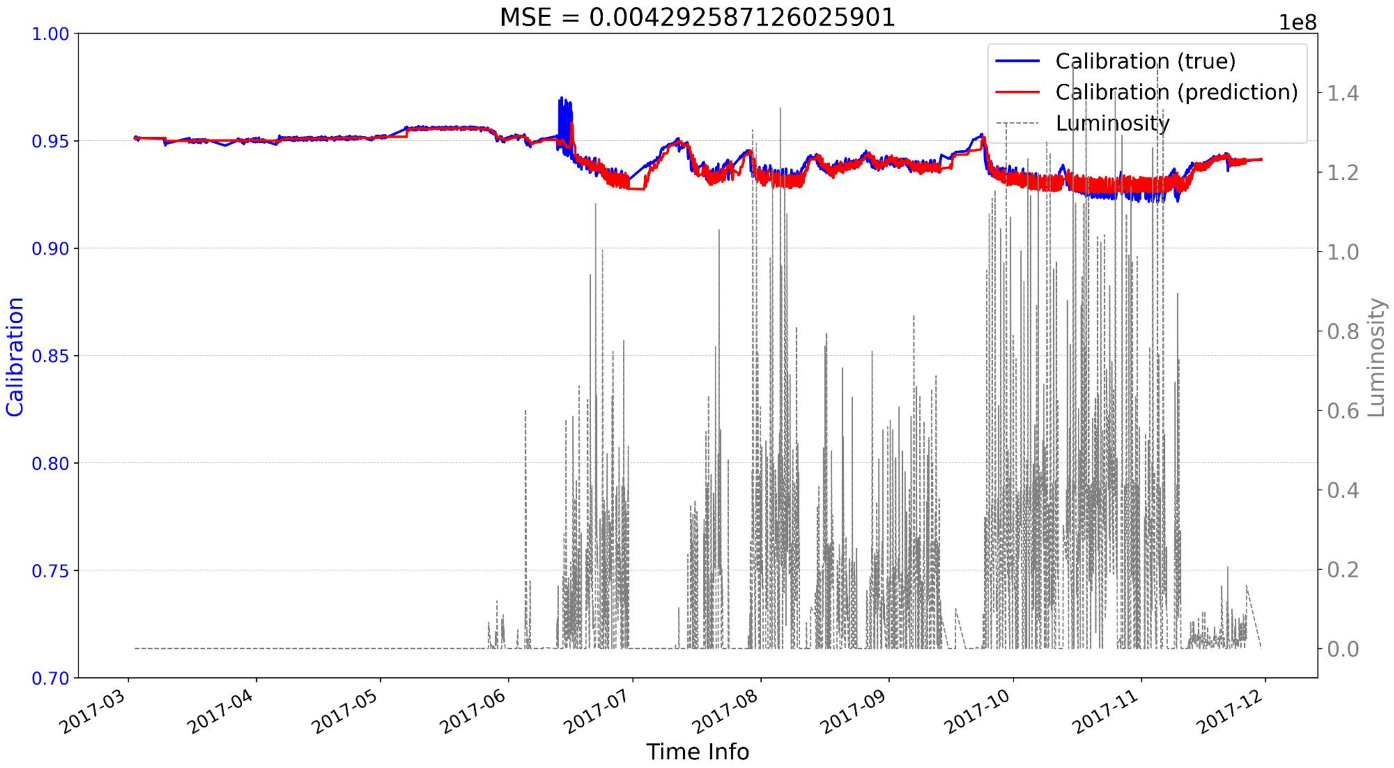



Year=2016; Mean=0.0; Std=1.0
No.
0
200
400
600
800
1000
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
Cali.
Year=2017; Mean=-0.0; Std=1.0
No.
0
200
400
600
800
1000
1200
1400
−2
−1
0
1
2
3
Cali.
Year=2018; Mean=0.0; Std=1.0
No.
0
200
400
600
800
1000
1200
1400
−2
−1
0
1
2
Cali.

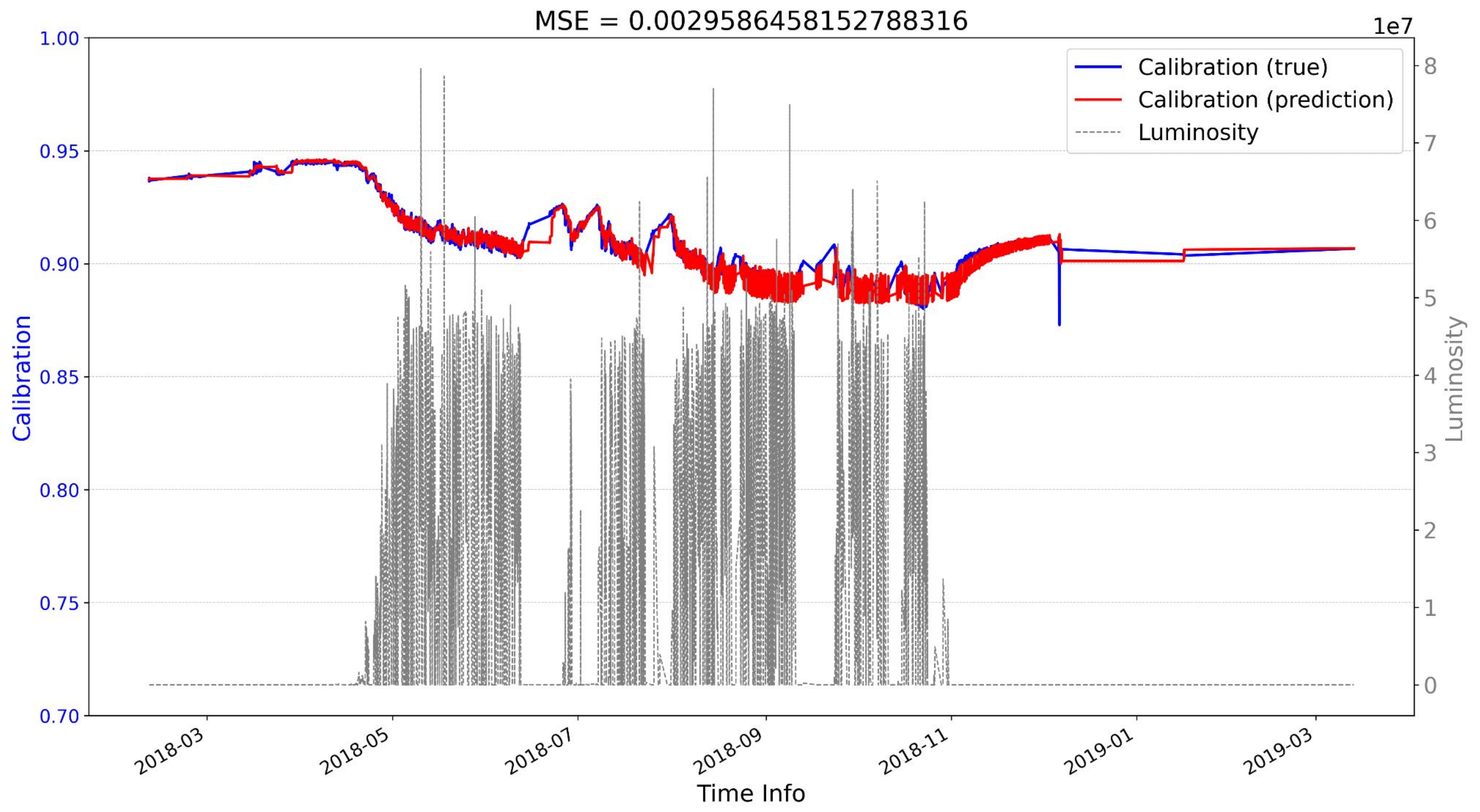
Results—Training 0n 2016; Test on 2017 & 2018



Results—Training 0n 2016; Test on 2017 & 2018



Results—Training 0n 2016; Test on 2017 & 2018



What’s Next

1) Test the model on more different crystals and years;
2) Try Case 2 (see slide 7 for details);
3) Add the results of model-type-1 (see slide 6 for details);
4) ……

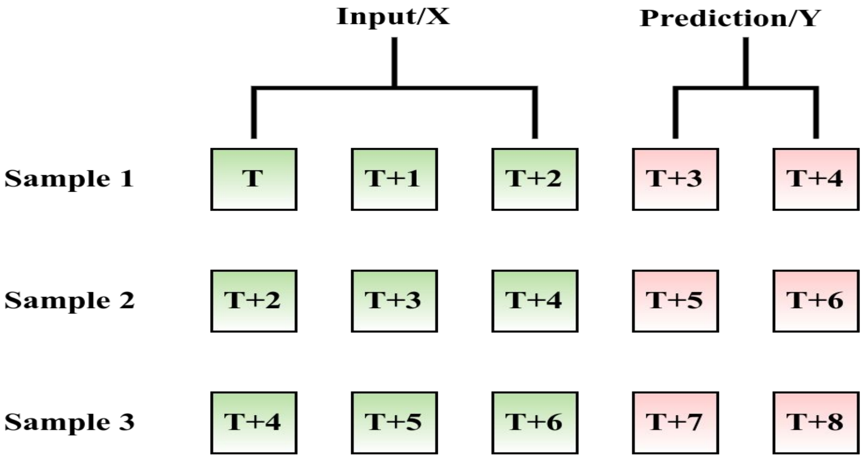
Training/Test Data Format

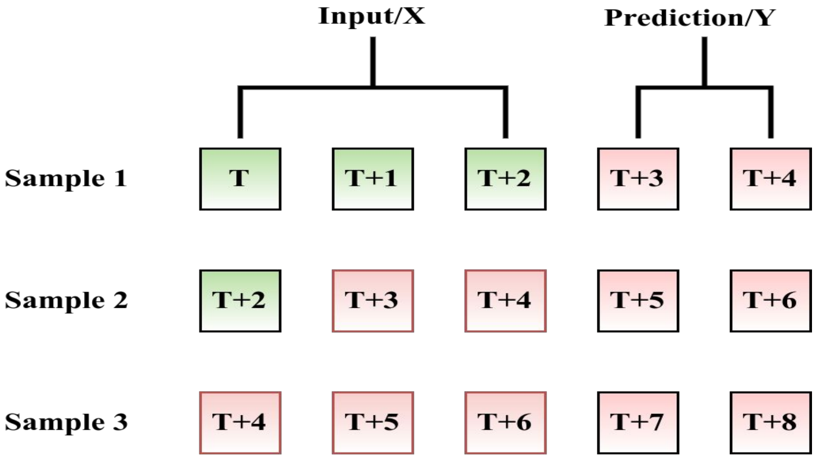
Case1 (left):

1) We can always observe 3 consecutive actual values and then make predict on the next two values;
2) When we predict "T+3 & T+4", we use the actual "T, T+1, T+2";
3) When we want to predict "T+5 & T+6", we wait until we obtained the actual "T+3 & T+4".

Case 2 (right):

1) The only observed information we have is "T, T+1, T+2";
2) In order to make much further prediction, we need to "re-use" our prediction as "fake observation".

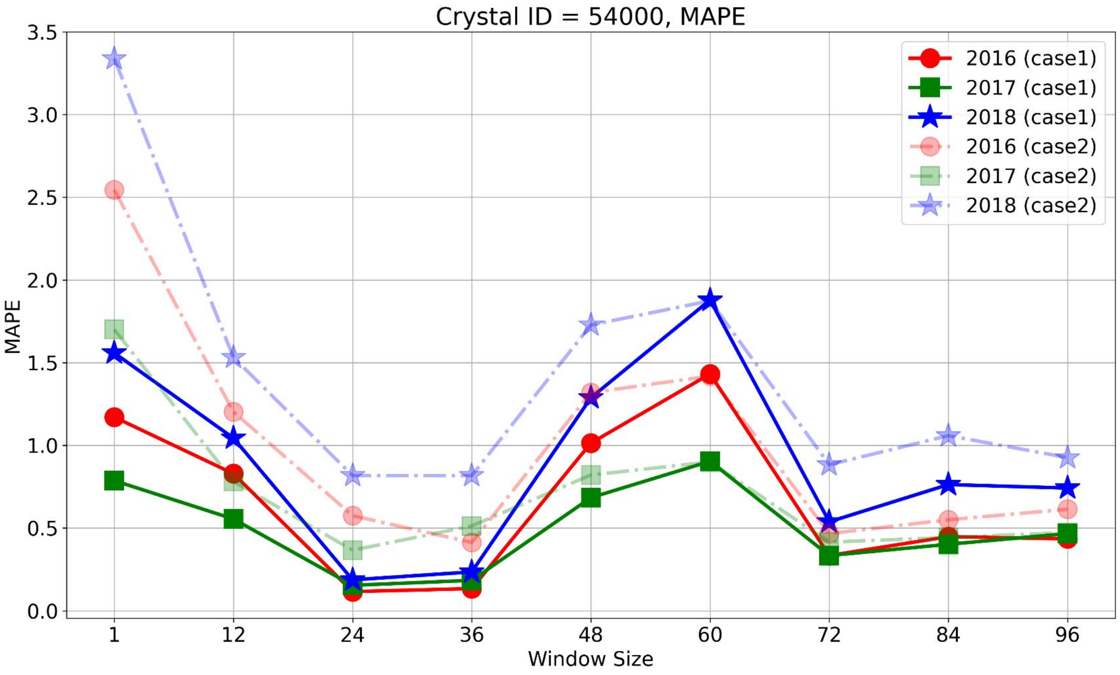




Crystal ID=54000, Different Window Size

Mean Absolute Percent Error (MAPE):
the lower, the better.

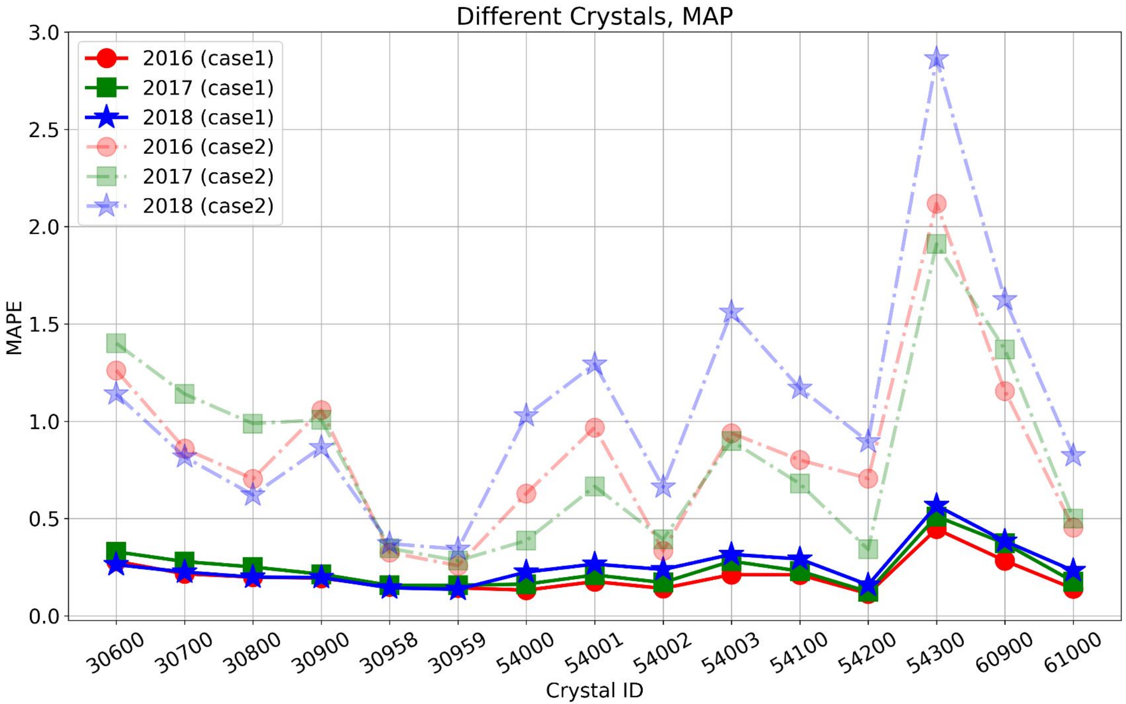
$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n}$$



Different Crystals, WS=24, Trained on 2016 (separately)

Mean Absolute Percent Error (MAPE):
the lower, the better.

$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n}$$



Different Crystals, WS= 24, Trained on 2016 (ID:54000)

Mean Absolute Percent Error (MAPE):
the lower, the better.

$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n}$$