Using Neural Networks to Predict Radiation Damage to Lead Tungstate Crystals at the CERN LHC

Taihui Li & Buyun Liang

University of Minnesota
Driven to Discover℠

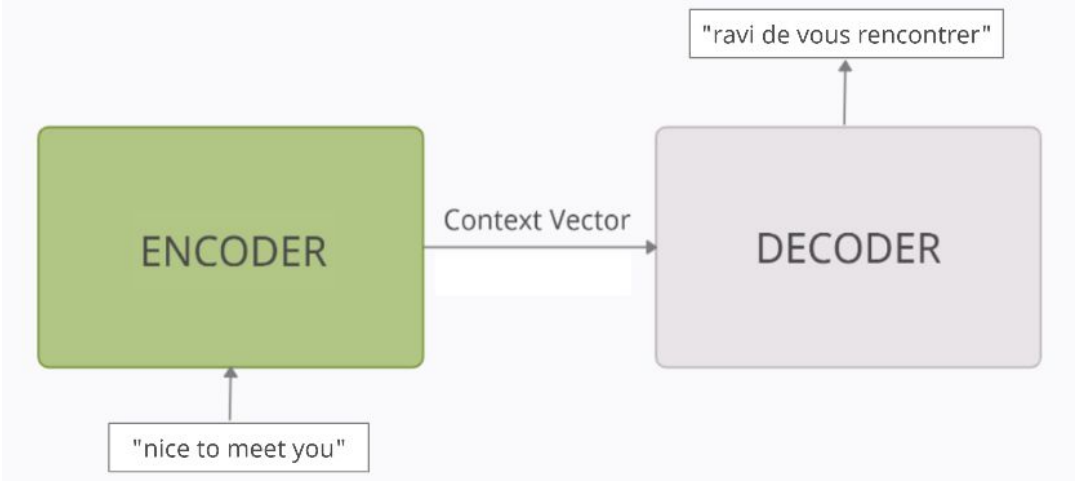
Seq2Seq Model

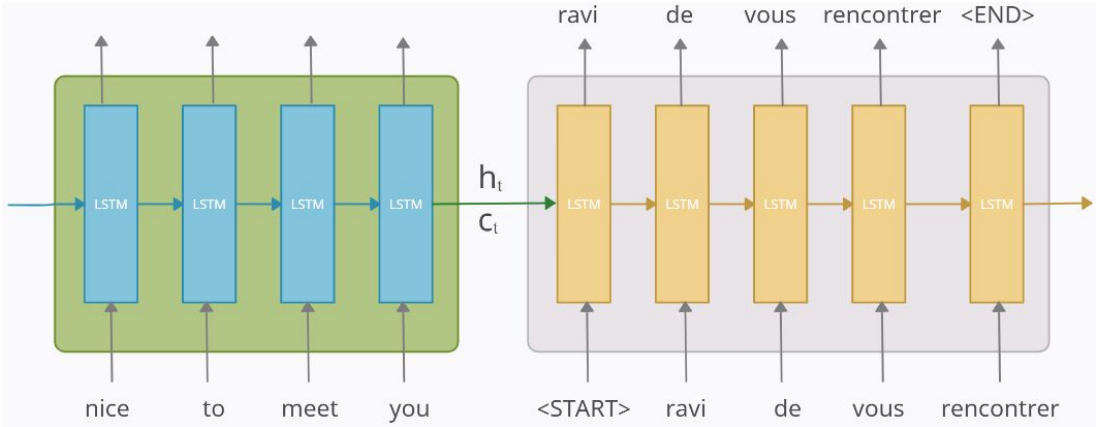
Input: (Enighlish) “nice to meet you”
Output: (French) “ravi de vous rencontrer”

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.

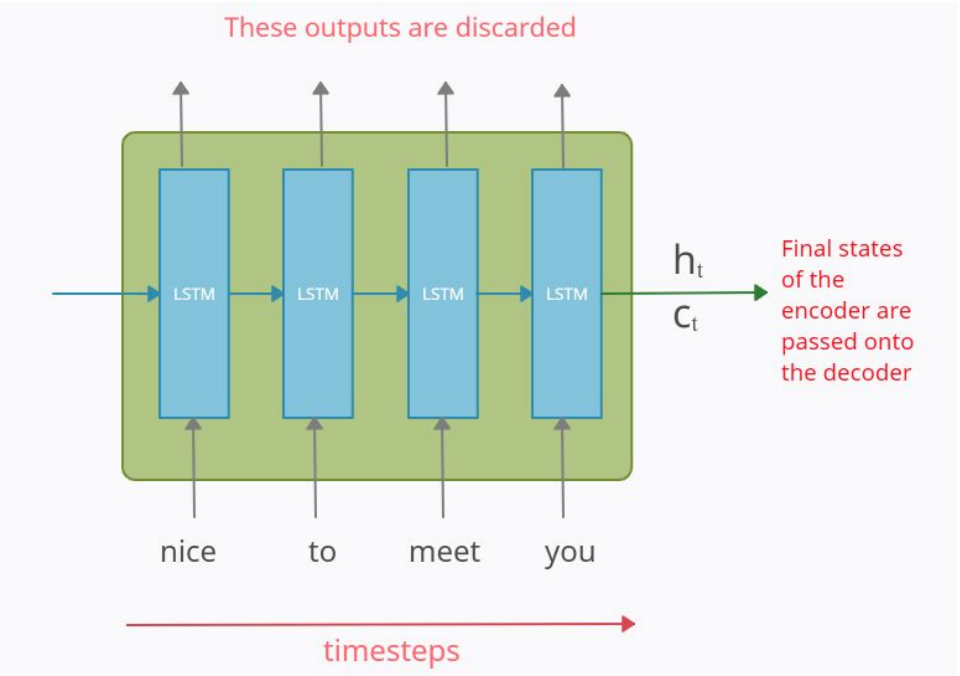
Decoder: Reading the context vector and tries to predict the target-seq token by token.

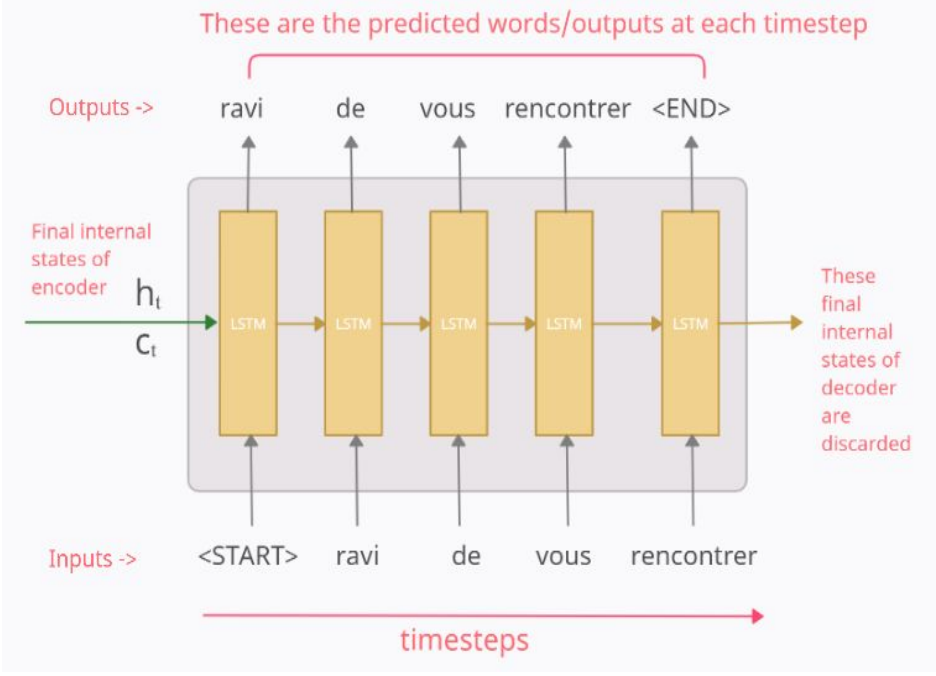




Ref:
https://medium.com/analytics-vidhya/encoder-decoder-seq2seq-models-clearly-explained-c34186fbf49b

Seq2Seq Model





Encoder **Decoder**

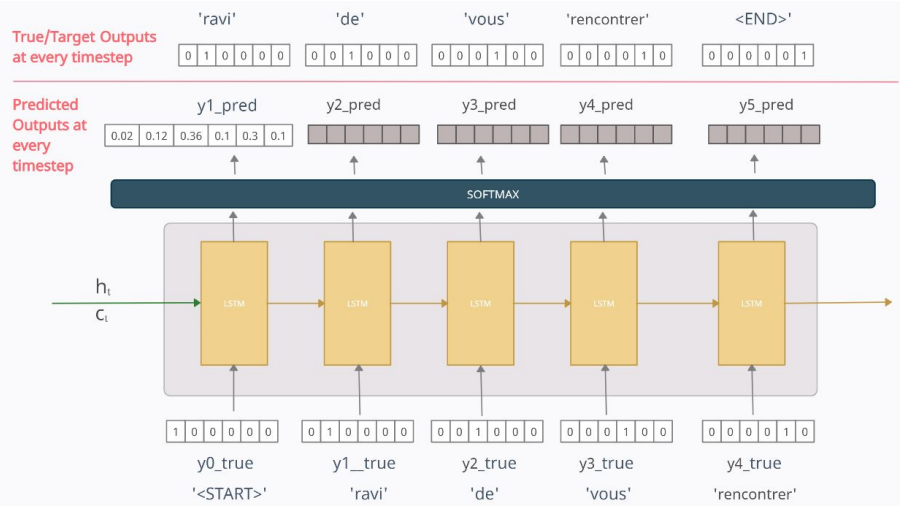
Ref:
https://medium.com/analytics-vidhya/encoder-decoder-seq2seq-models-clearly-explained-c34186fbf49b

Seq2Seq Training & Test

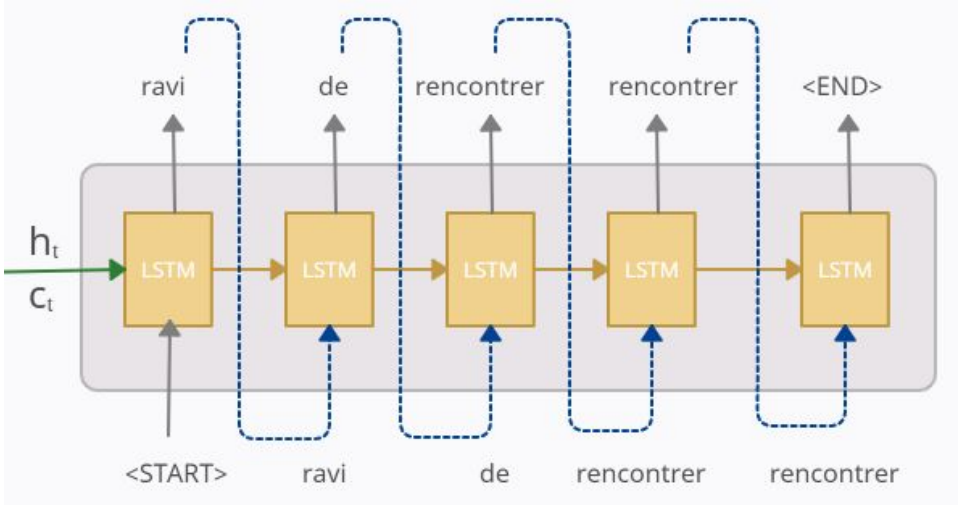
The Decoder in Training Phase:

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

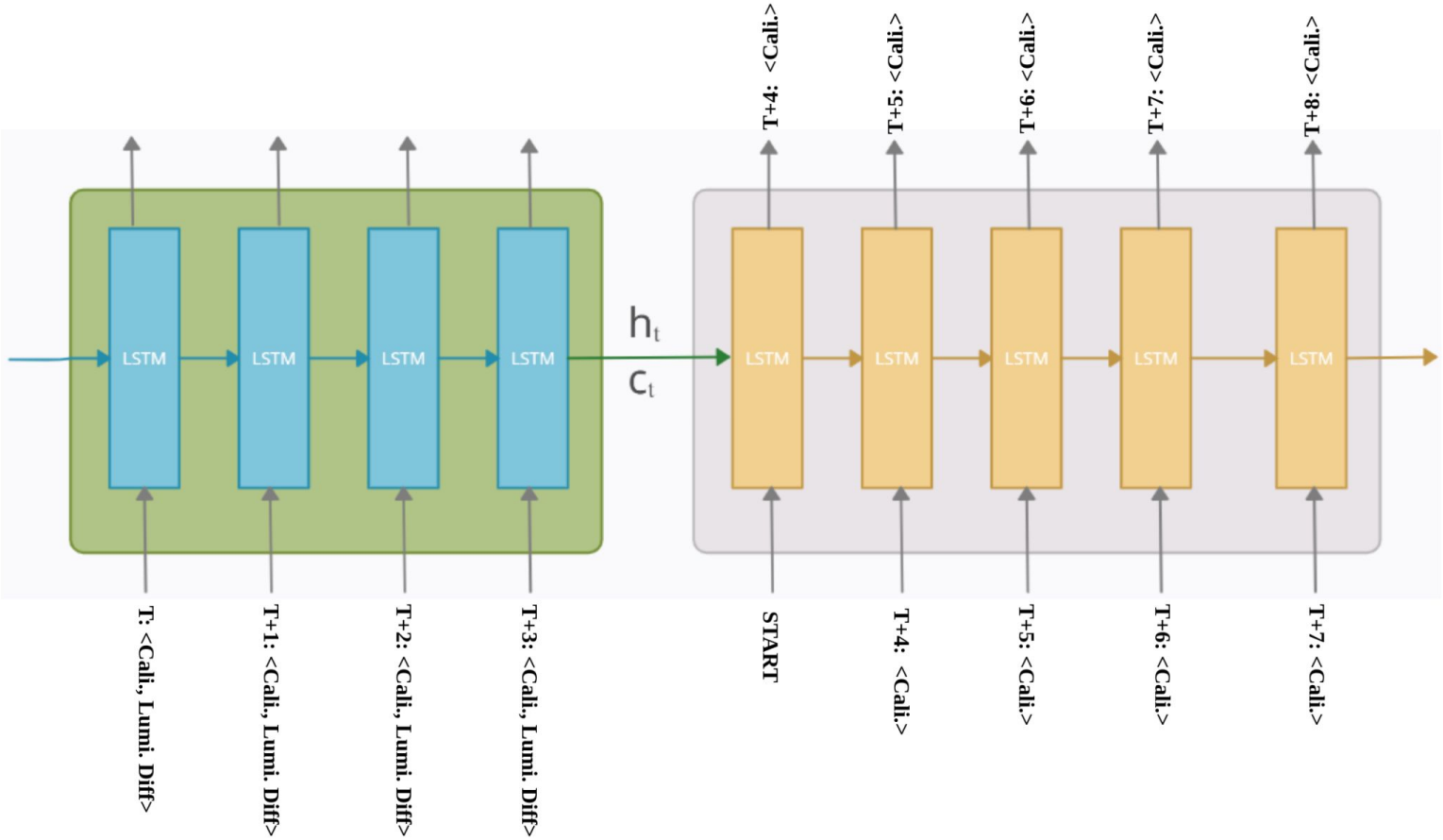


Teacher Forcing

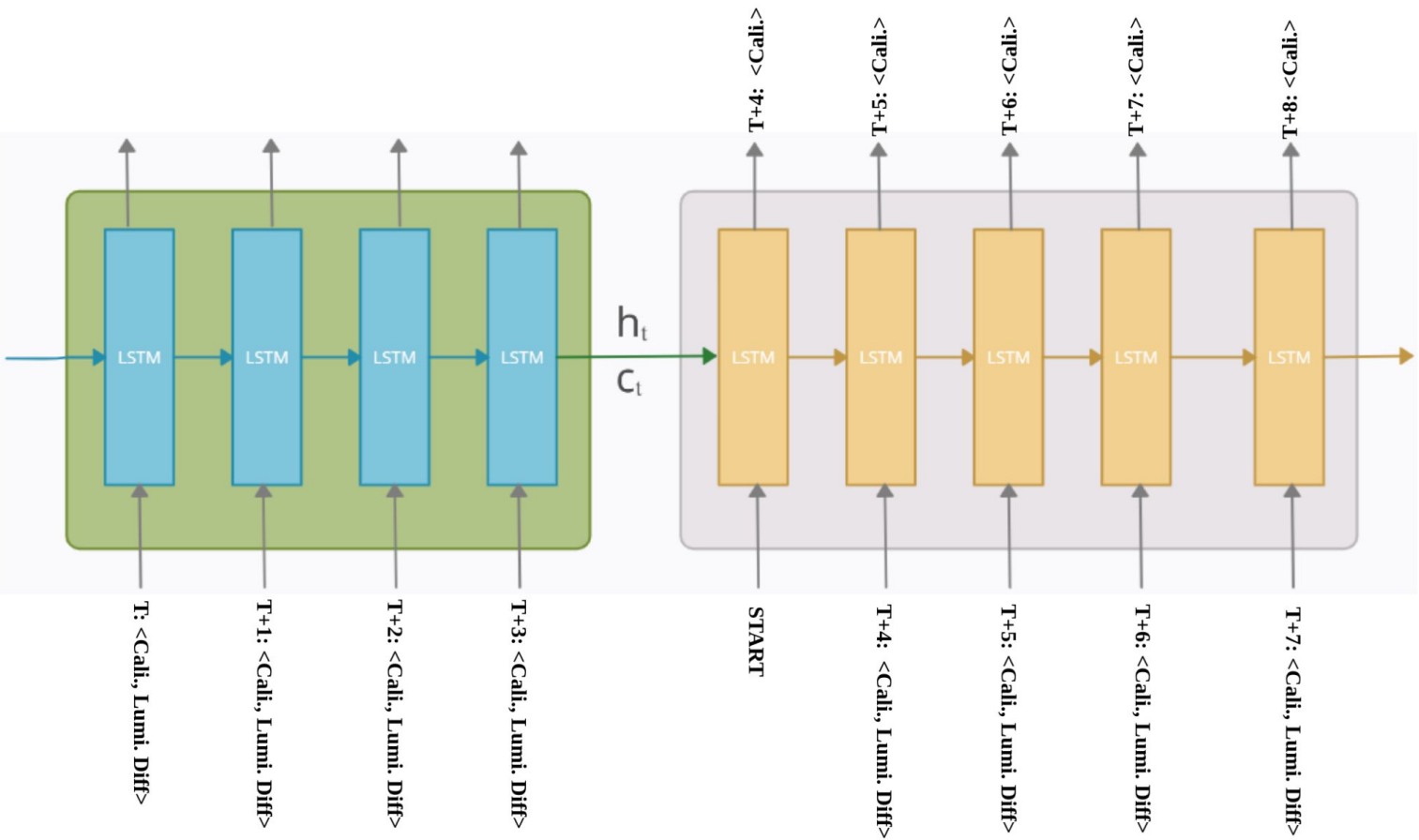


Without Teacher Forcing

Our Seq2Seq Model Type-1



Our Seq2Seq Model Type-2



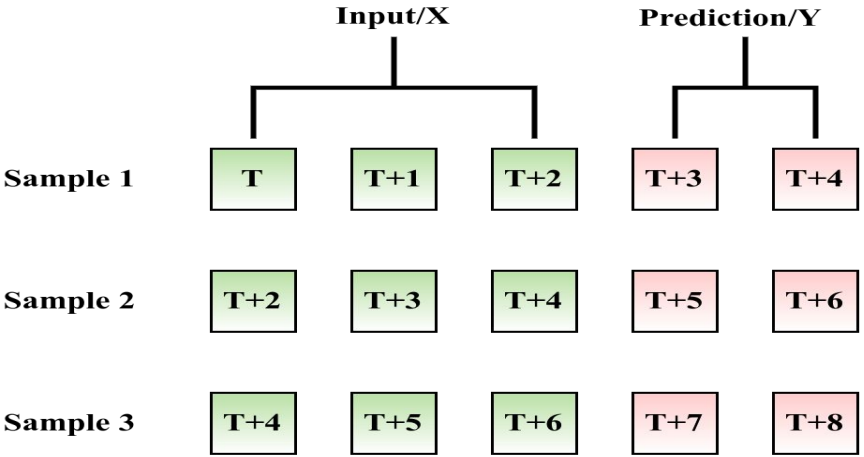
Training/Test Data Format

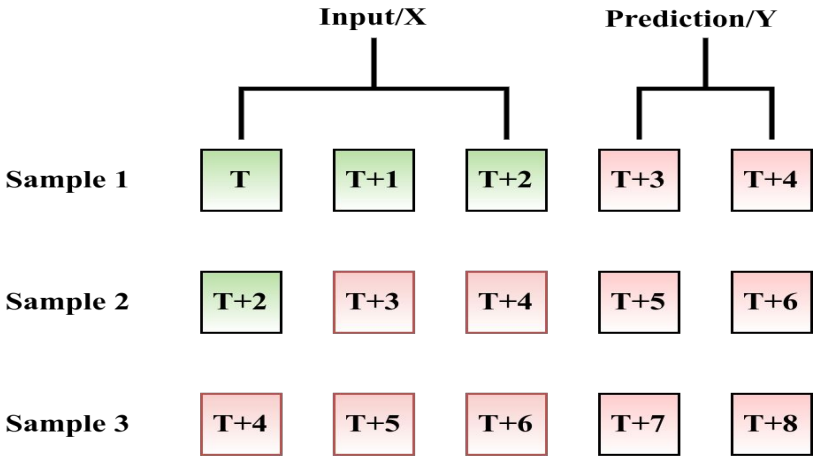
Case1 (left):

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".



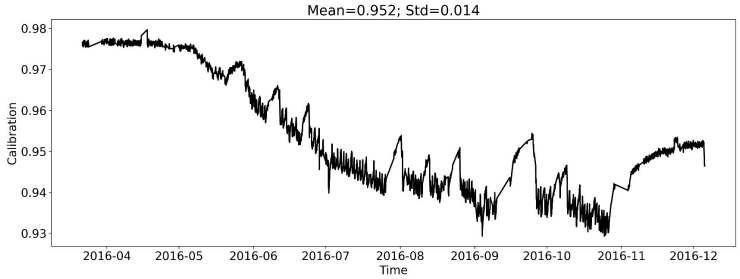


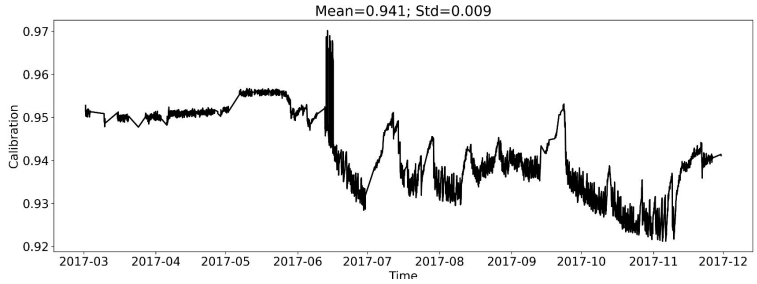
Experimental Setting Up

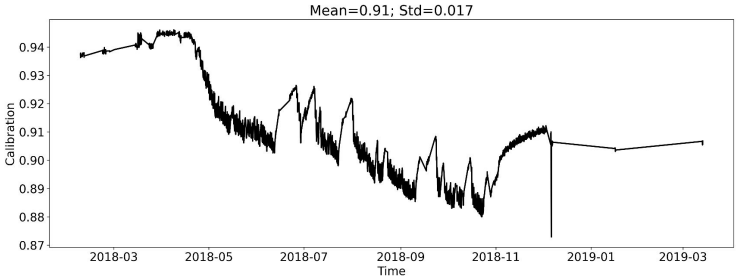
All results in the following slides use the same setting up:

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, 2018 data of 54000 crystal.

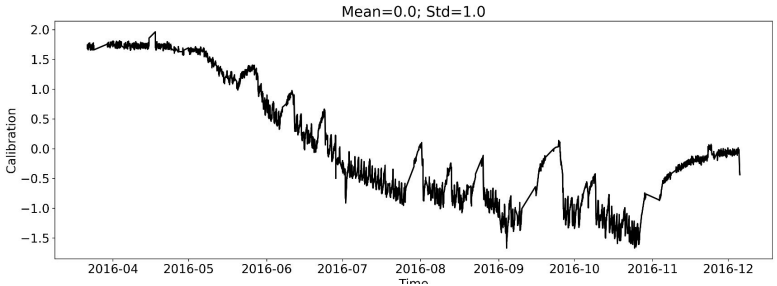
Original Calibration

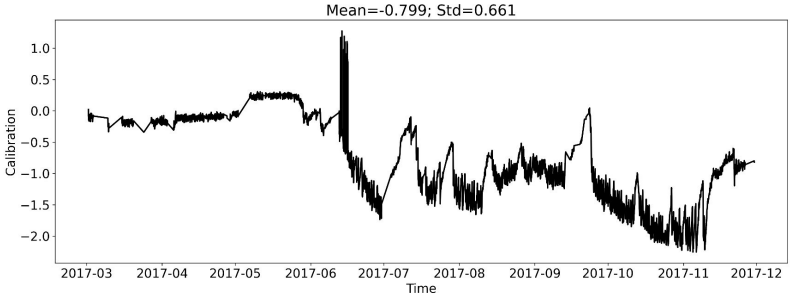


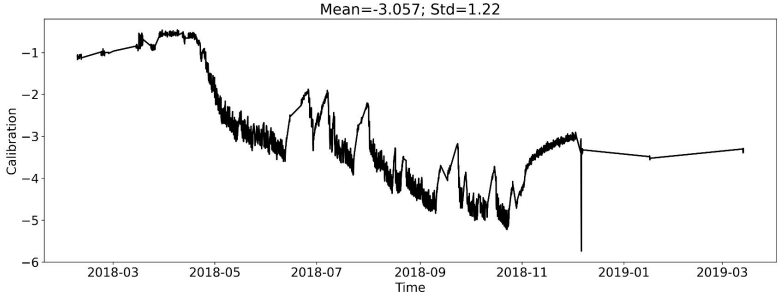




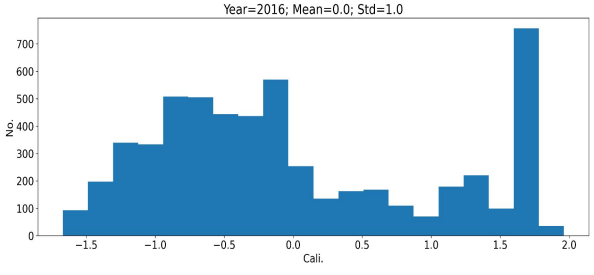
Normalized Calibration

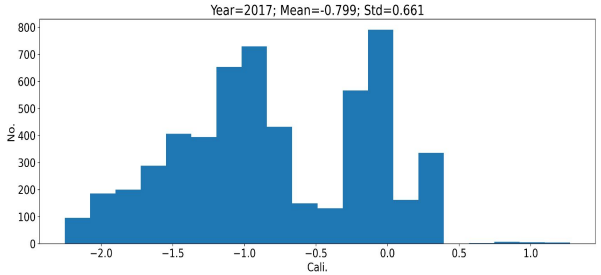


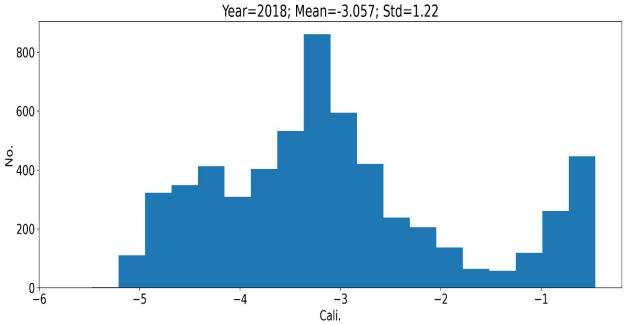




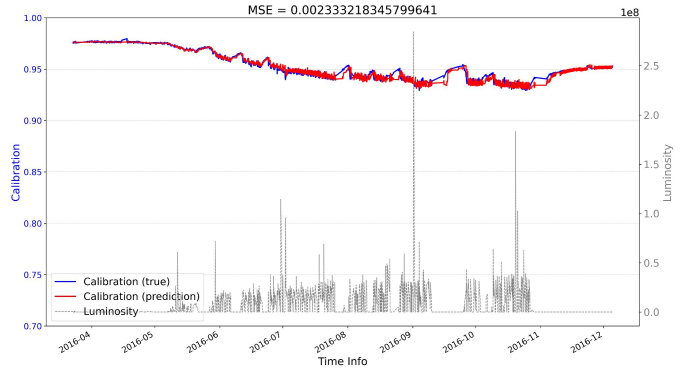
Calibration Distribution

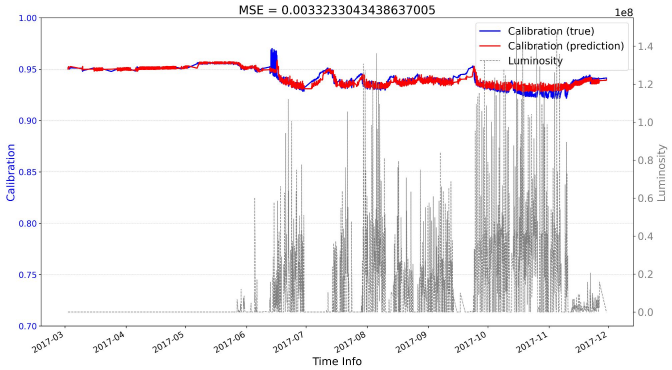


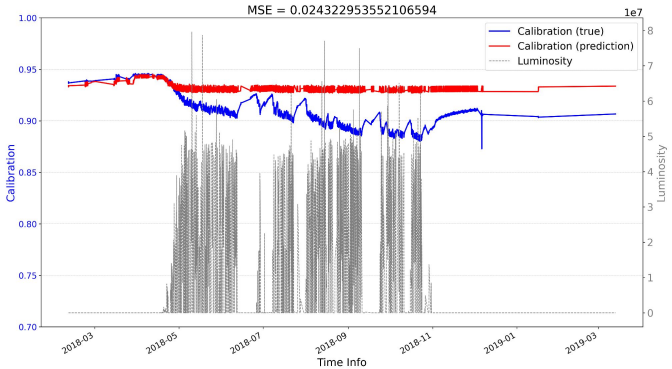




Results—Training on 2016; Test on 2017 & 2018

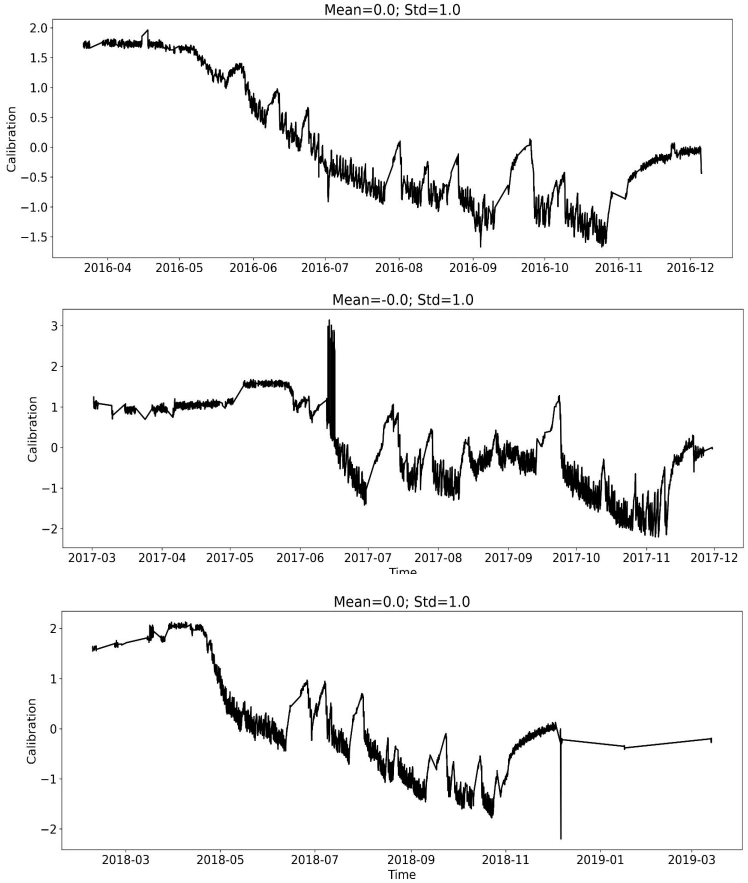




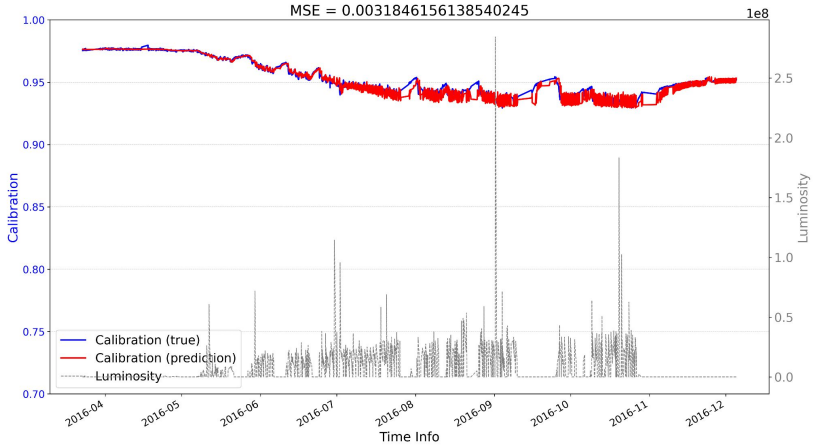


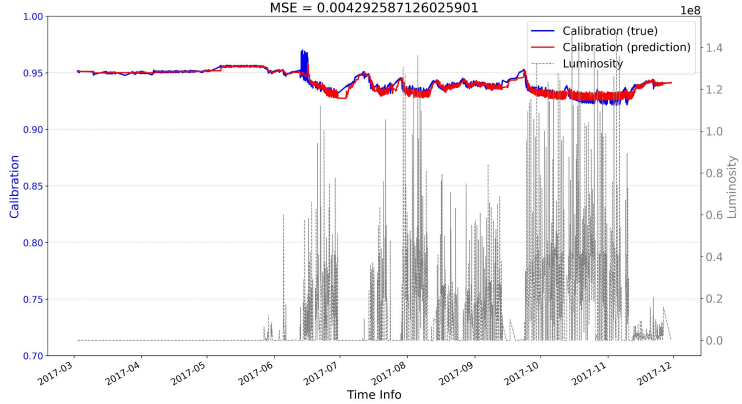
Data distribution shift causes the prediction performance degradation

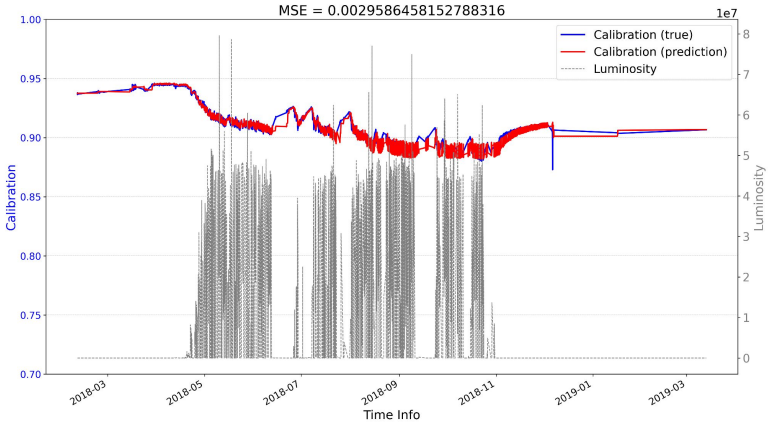
Normalize the data separately



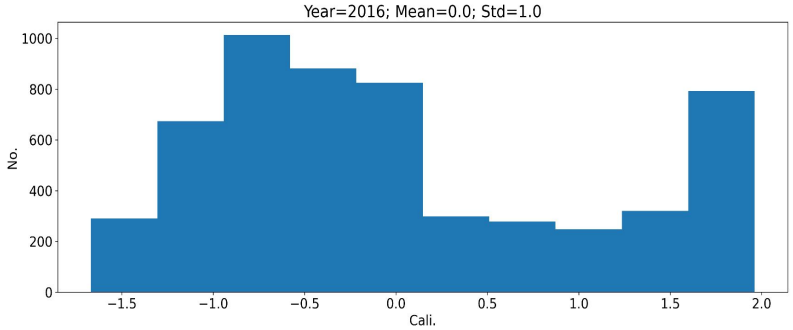
Results—Training on 2016; Test on 2017 & 2018

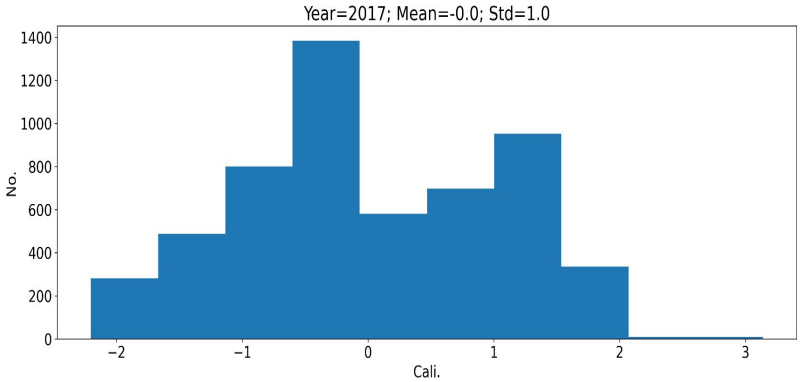


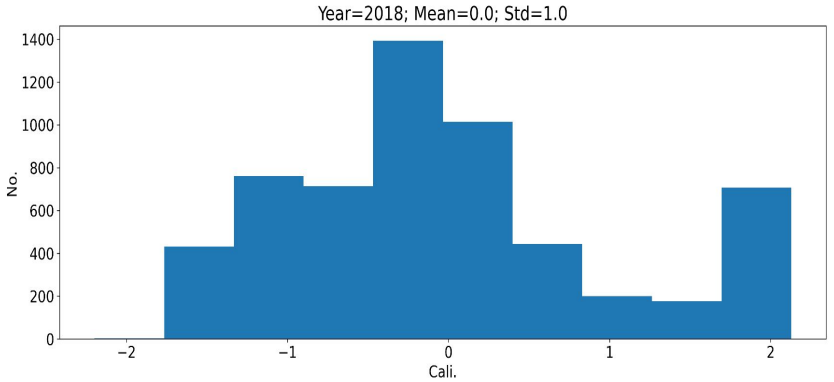




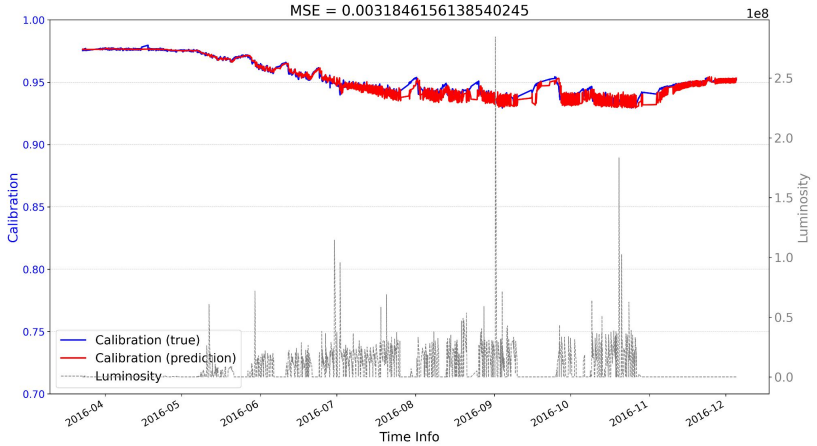
Calibration Distribution

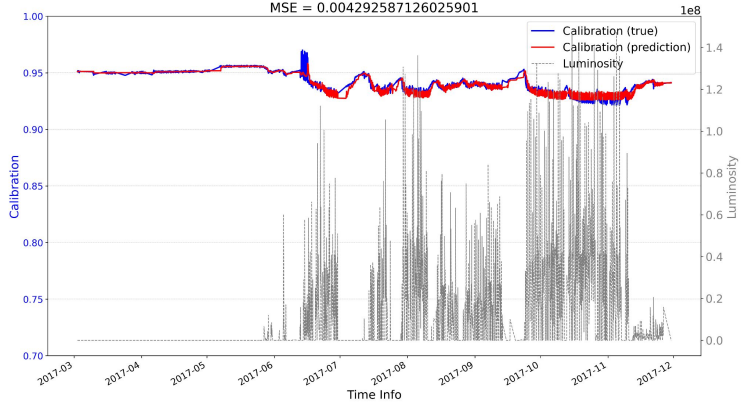


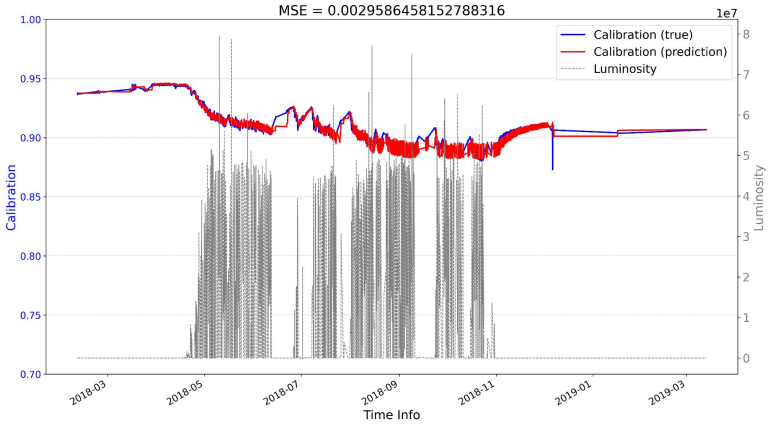




Results—Training on 2016; Test on 2017 & 2018



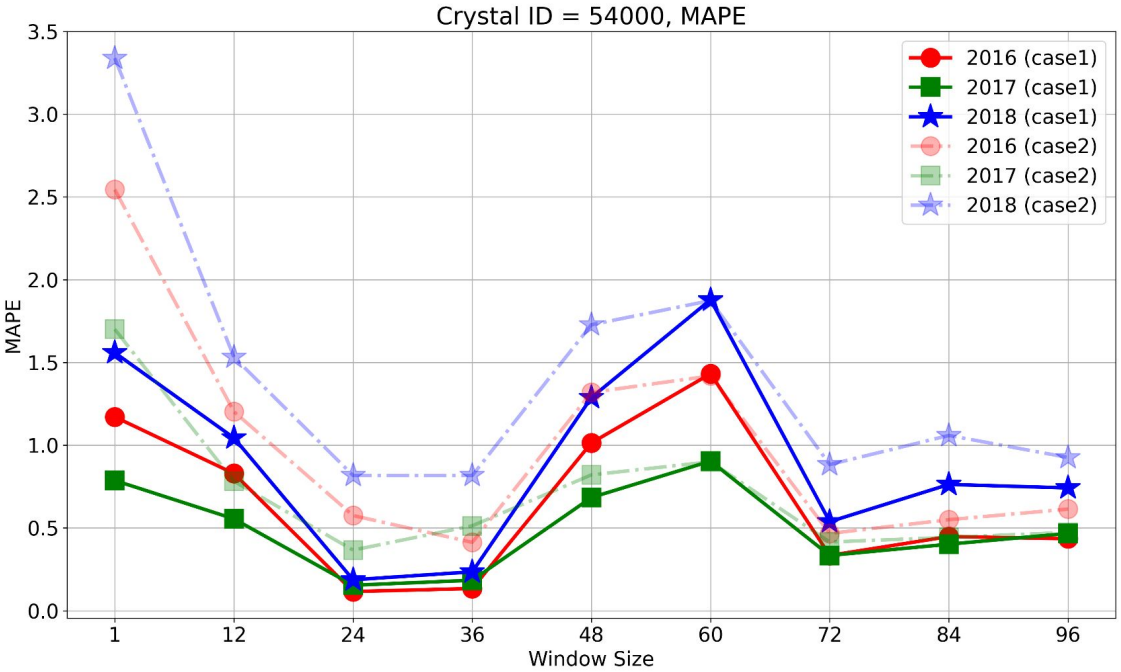




Crystal ID=54000, Different Window Size

Mean Absolute Percent Error (MAPE):
the lower, the better.
Window size = 24 gives the best performance

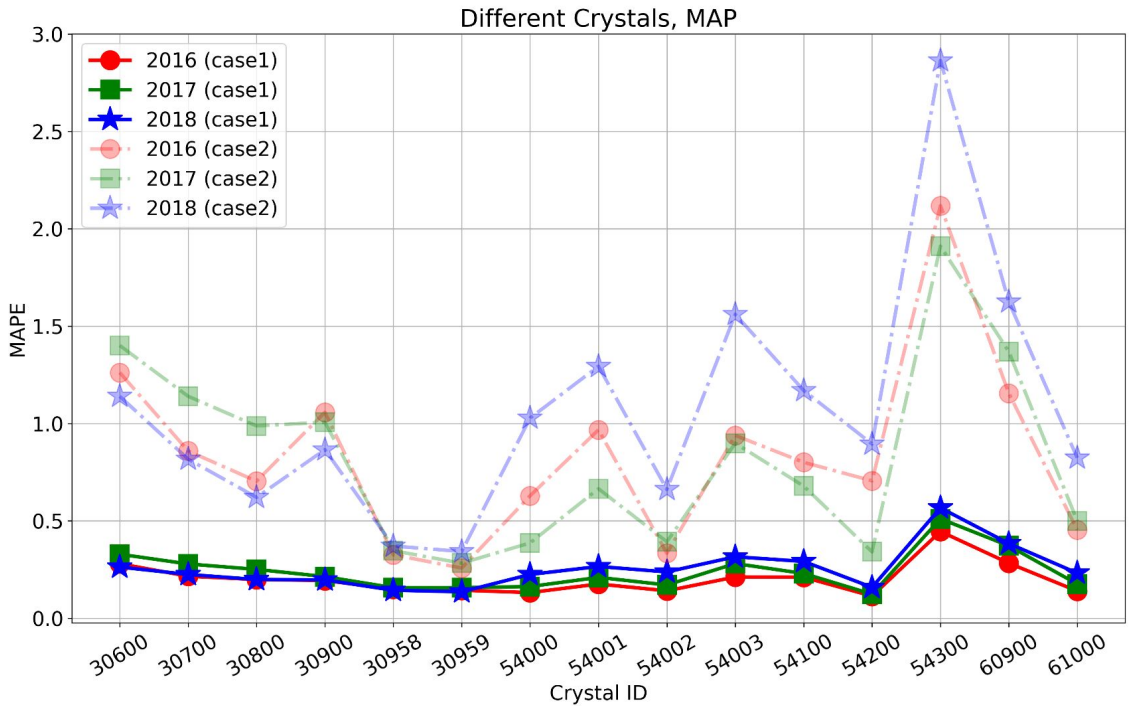
$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n}$$



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

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

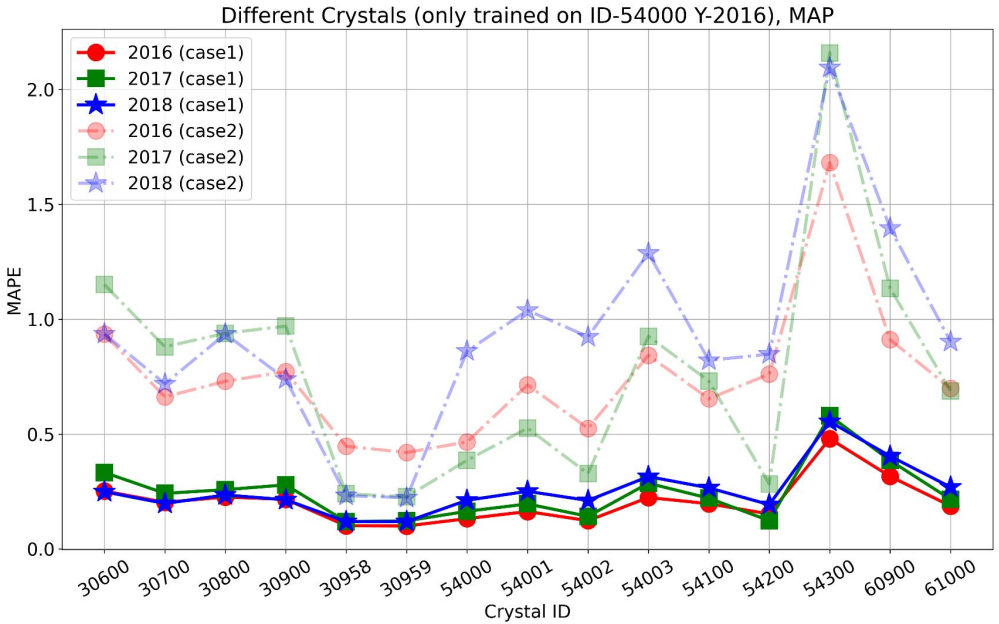
$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n}$$



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

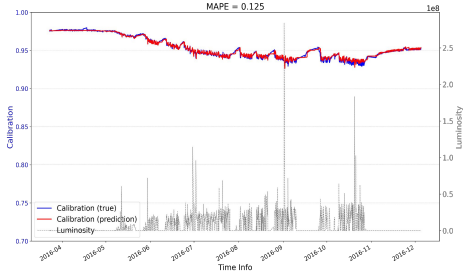
Mean Absolute Percent Error (MAPE):
the lower, the better.

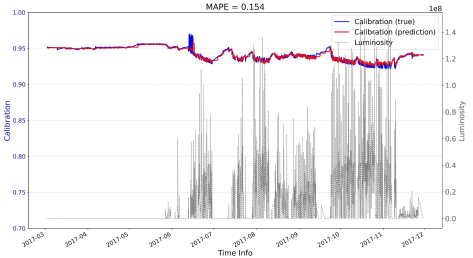
$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n}$$

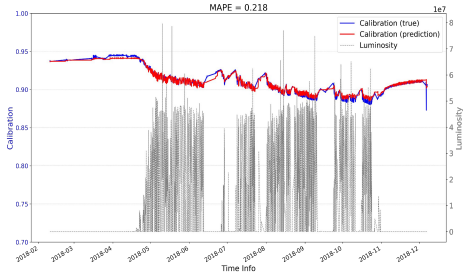


Different Test Strategies

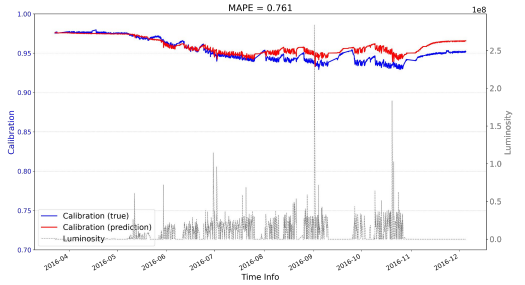
Case 1



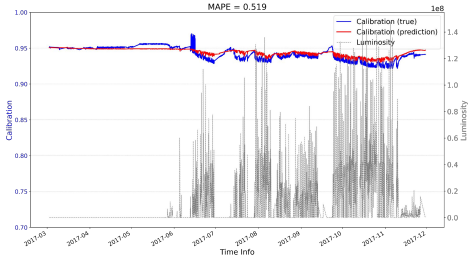


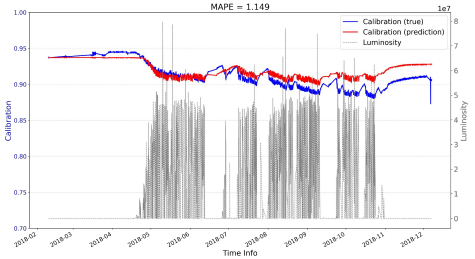


Trained on 2016 (ID:54000), predicted on 54000, mixed mode: teacher forcing ratio = 0.5



Case 2

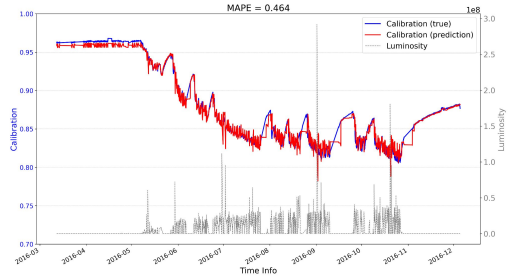


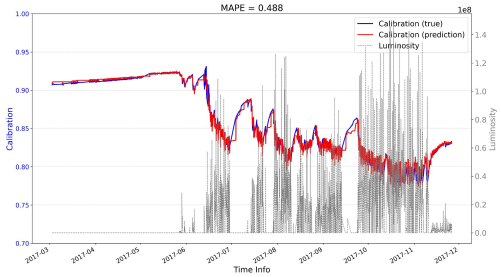


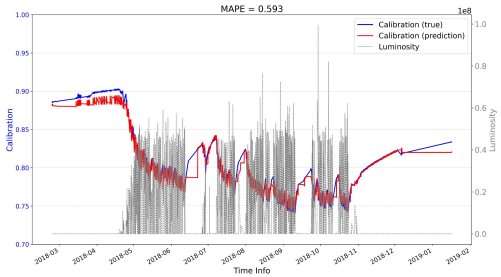
Different Test Strategies

Trained on 2016 (ID:54000), predicted on 54300, mixed mode: teacher forcing ratio = 0.5

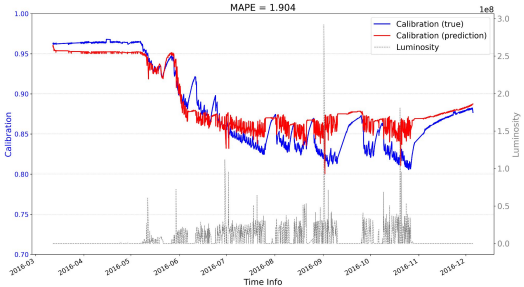
Case 1

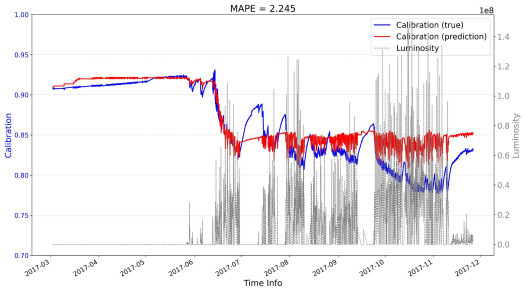


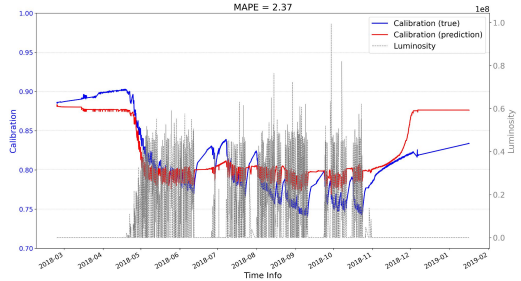




Case 2

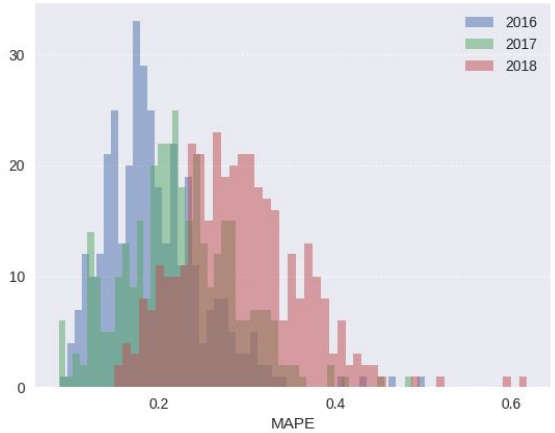




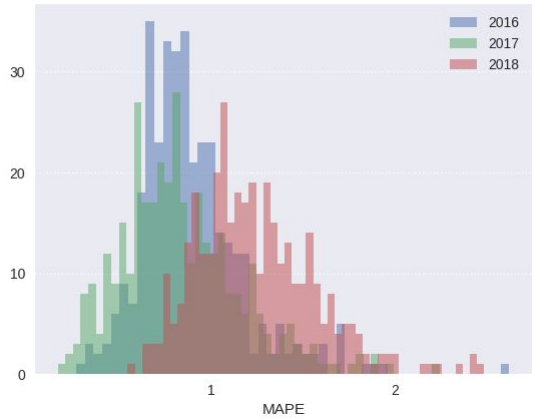


MAPE Histogram of Different Test Strategies

Trained on 2016 (ID:54000), predicted on 54000~54359,
mixed mode: teacher forcing ratio = 0.5



Case 1

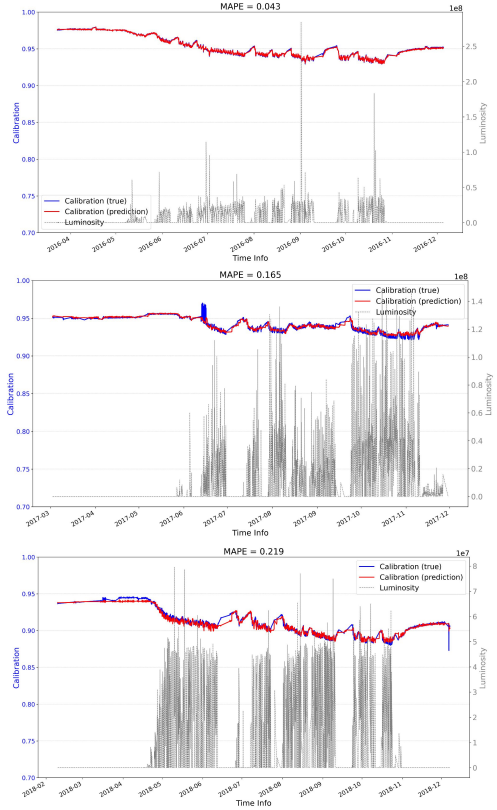


Case 2

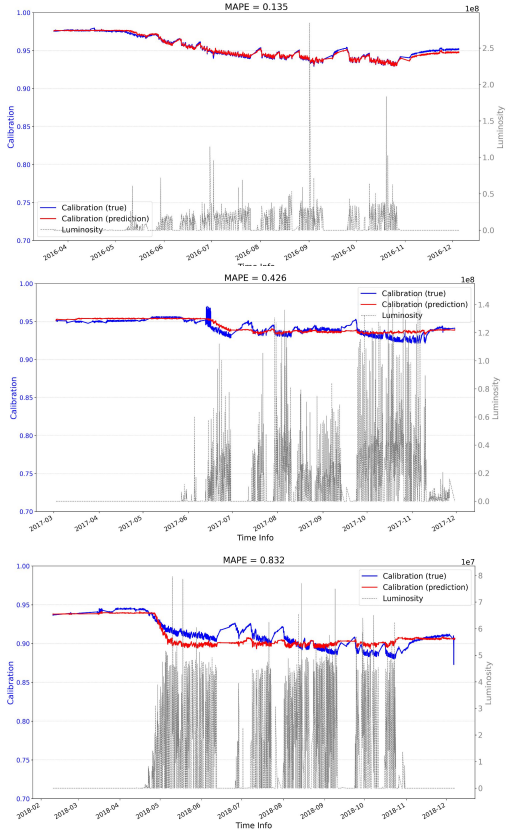
Trained on Multiple Crystals from a Ring

Trained on 2016 (ID:54000-54359), predicted on 54000, mixed mode: teacher forcing ratio = 0.5

Case 1



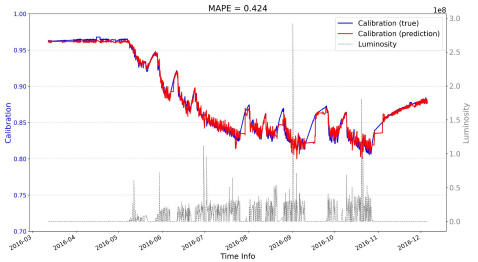
Case 2

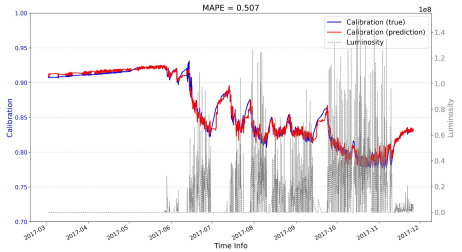


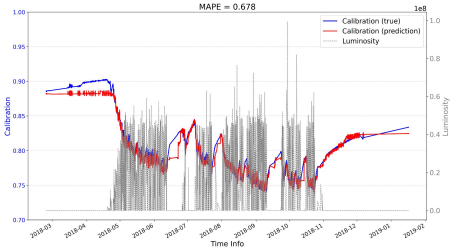
Trained on Multiple Crystals from a Ring

Trained on 2016 (ID:54000-54359), predicted on 54300, mixed mode: teacher forcing ratio = 0.5

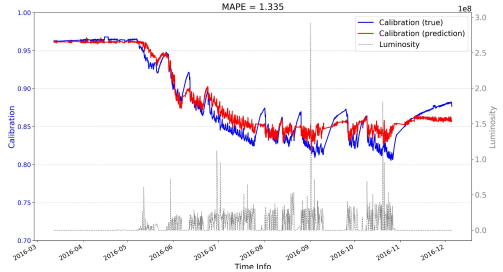
Case 1

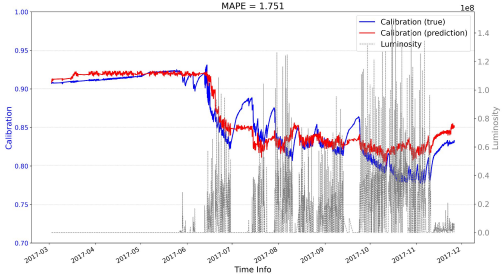


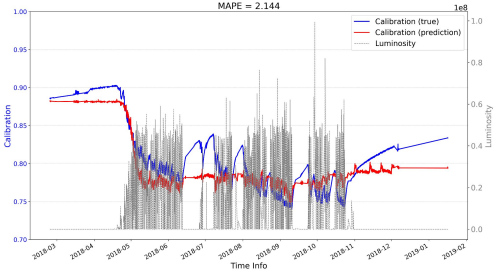




Case 2

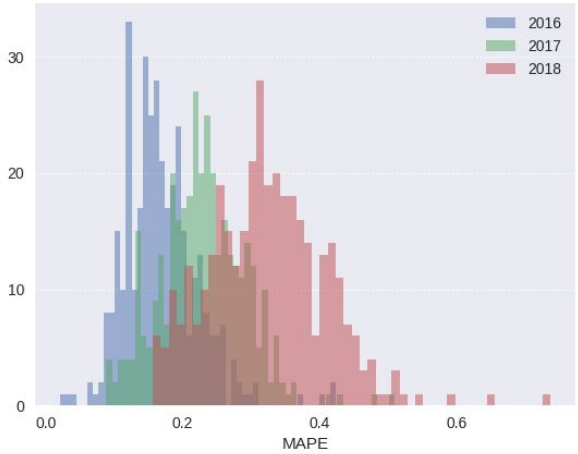




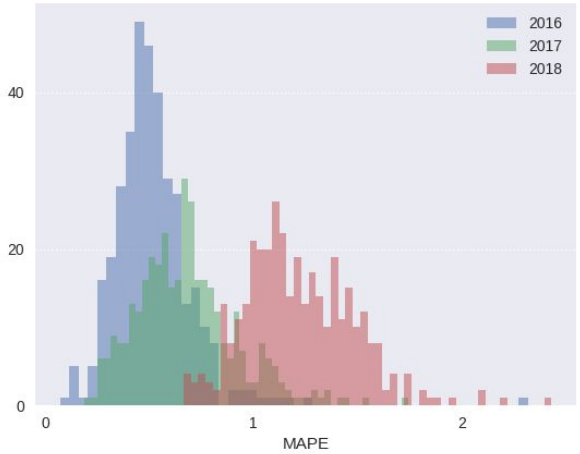


MAPE Histogram of Different Test Strategies

Trained on 2016 (ID:54000-54359), predicted on 54000~54359,
mixed mode: teacher forcing ratio = 0.5



Case 1

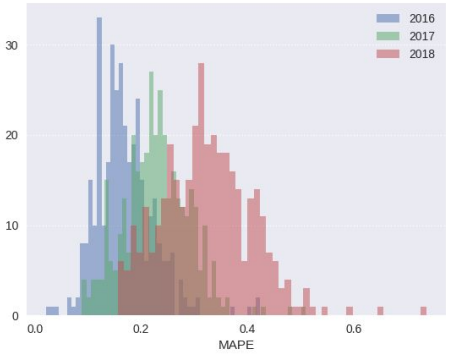


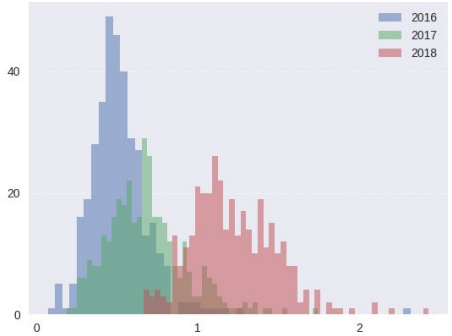
Case 2

MAPE Histogram of Different Training Strategies

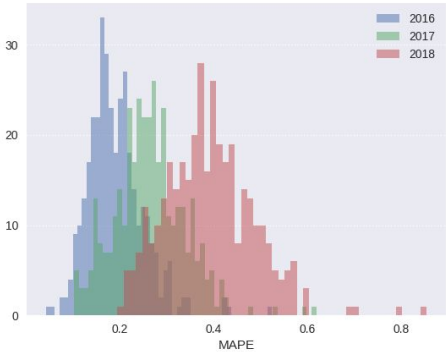
Trained on 2016 (ID:54000-54359), predicted on 54000~54359

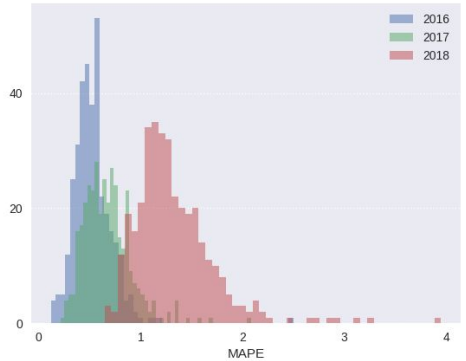
Recursive





Teacher forcing





Case 1

Case 2

MAPE of Different Training Strategies

Year	Prediction	Single	Ring (Recursive)	Ring (Mixed)	Ring (Teacher Forcing)
2016	Case 1	0.194	0.168	0.180	0.191
2017	Case 1	0.223	0.228	0.234	0.263
2018	Case 1	0.291	0.323	0.330	0.391
2016	Case 2	0.888	0.516	0.577	0.530
2017	Case 2	0.836	0.680	0.713	0.673
2018	Case 2	1.24	1.216	1.147	1.327