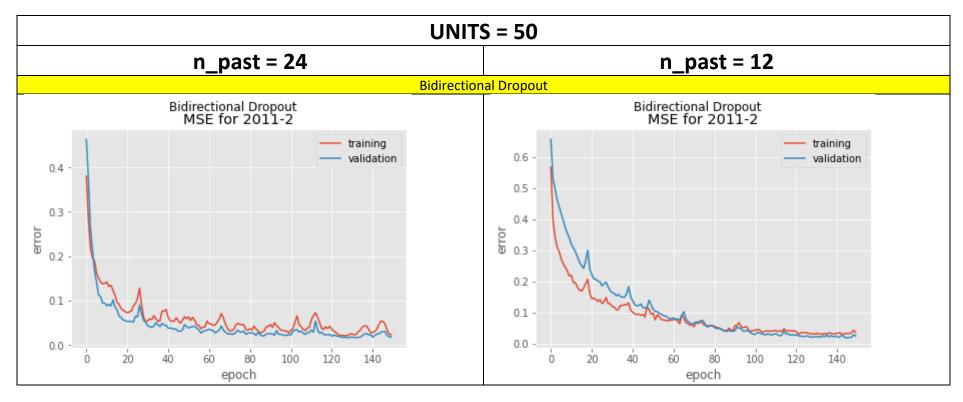
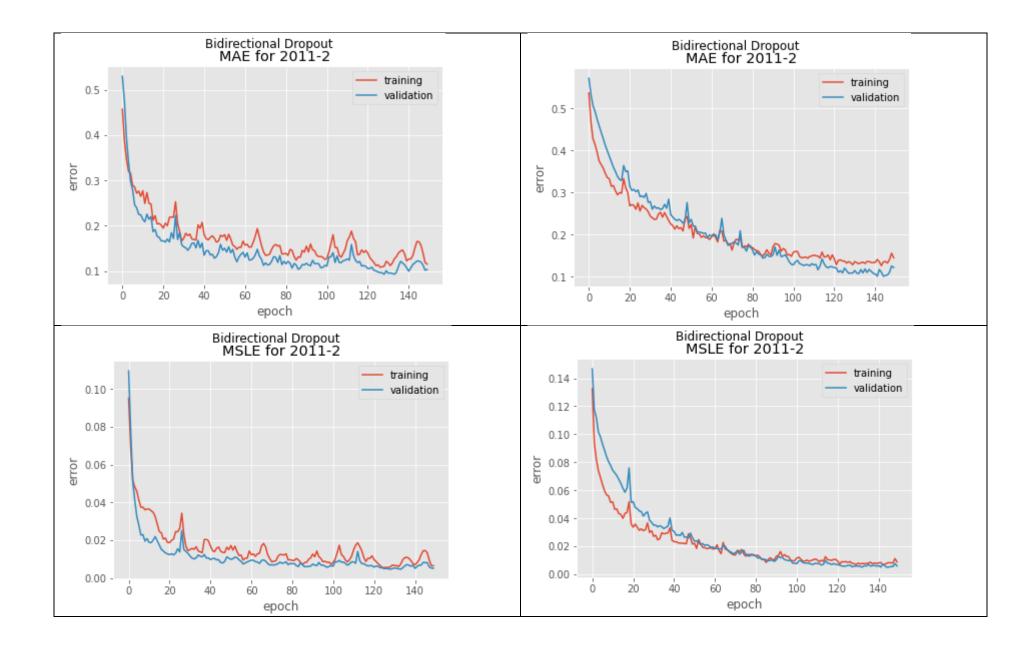
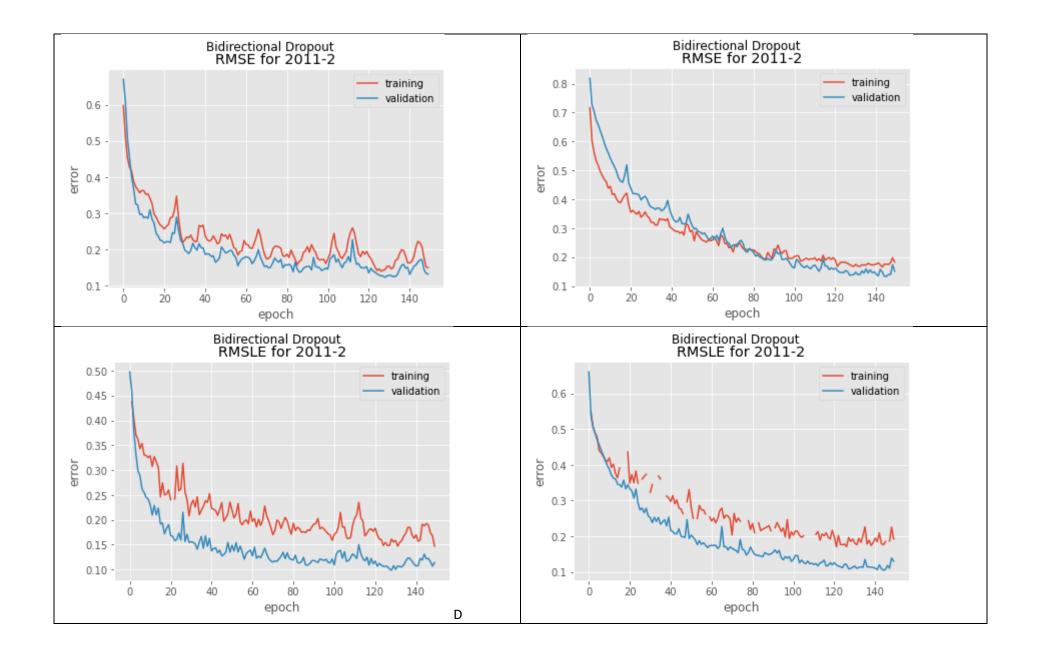
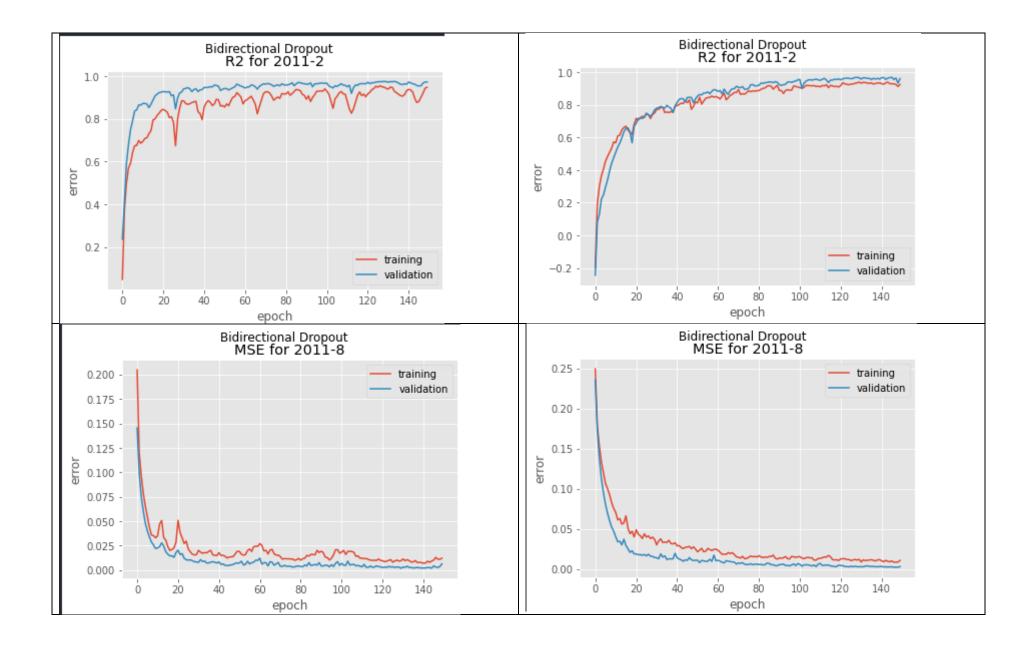
TRAINING HISTORY – CURVES

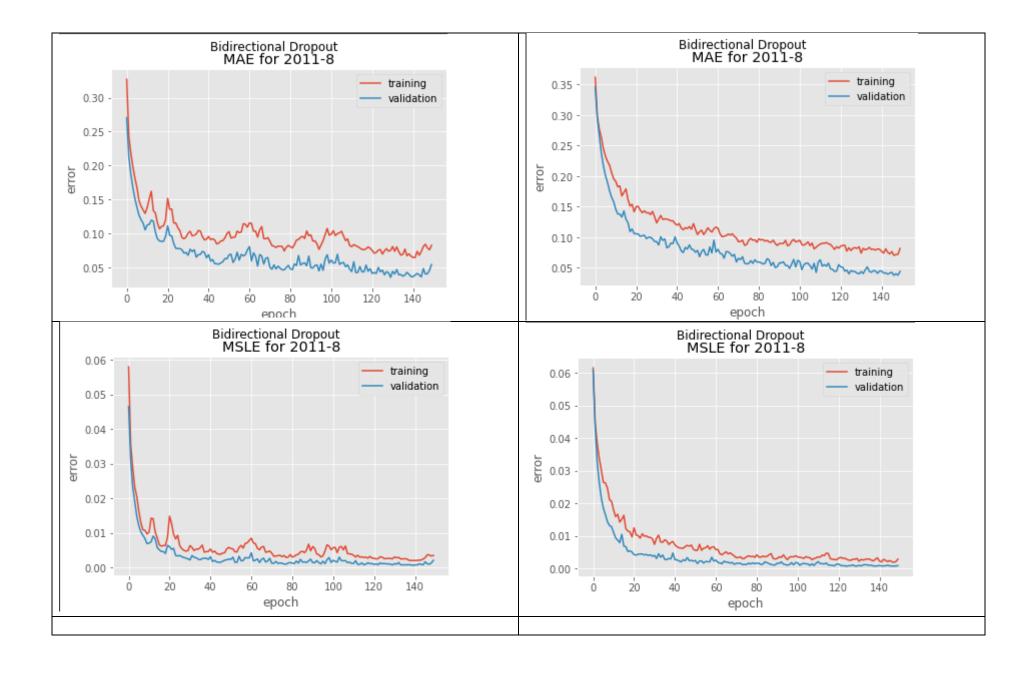
All curves are from models with linear activation function

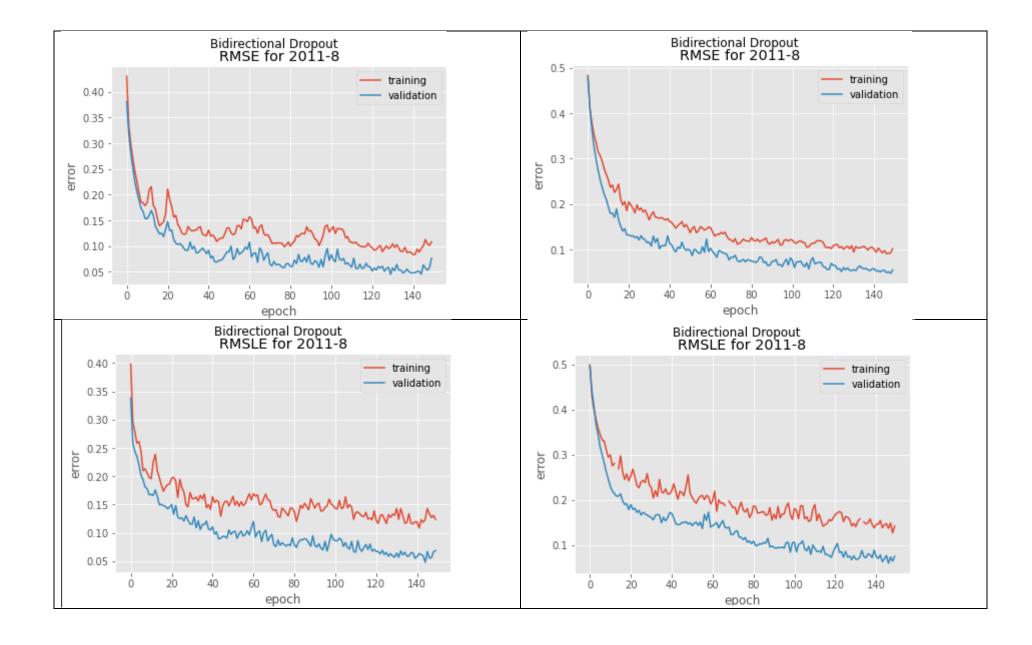


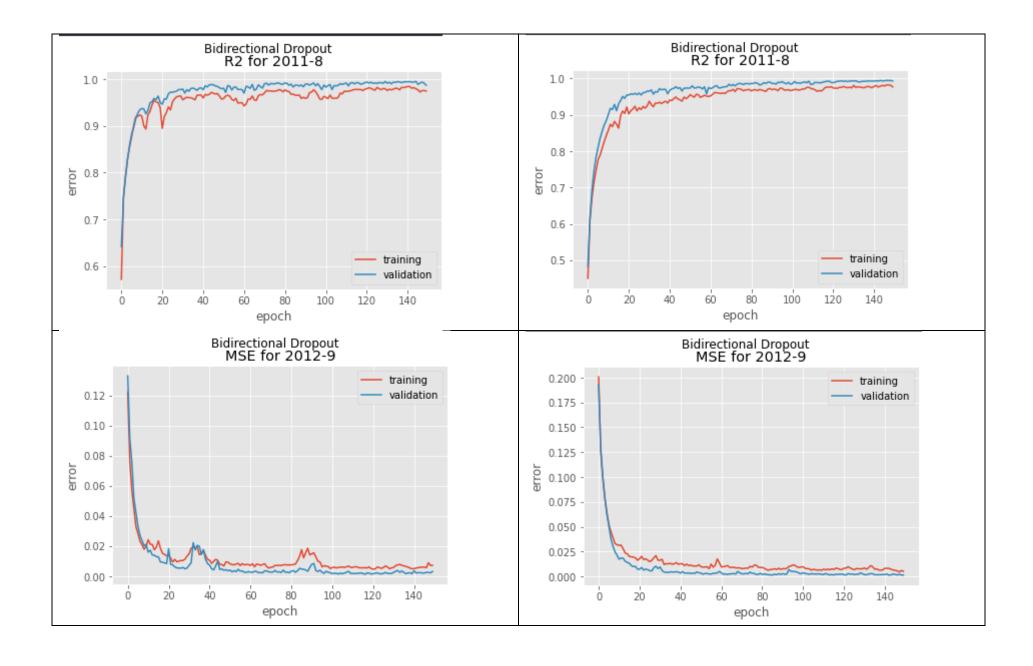


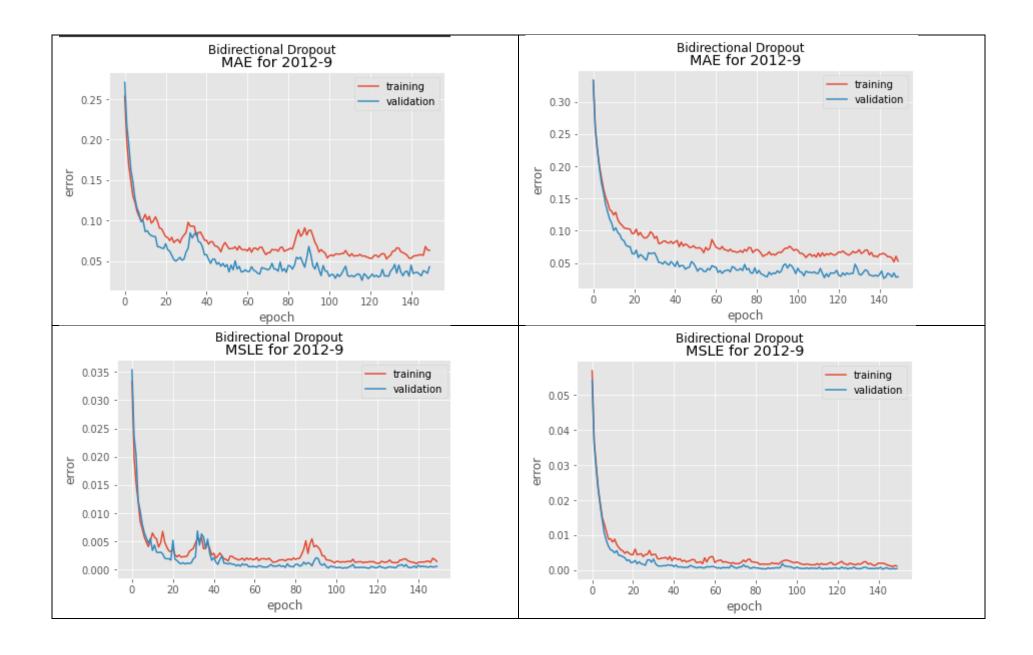


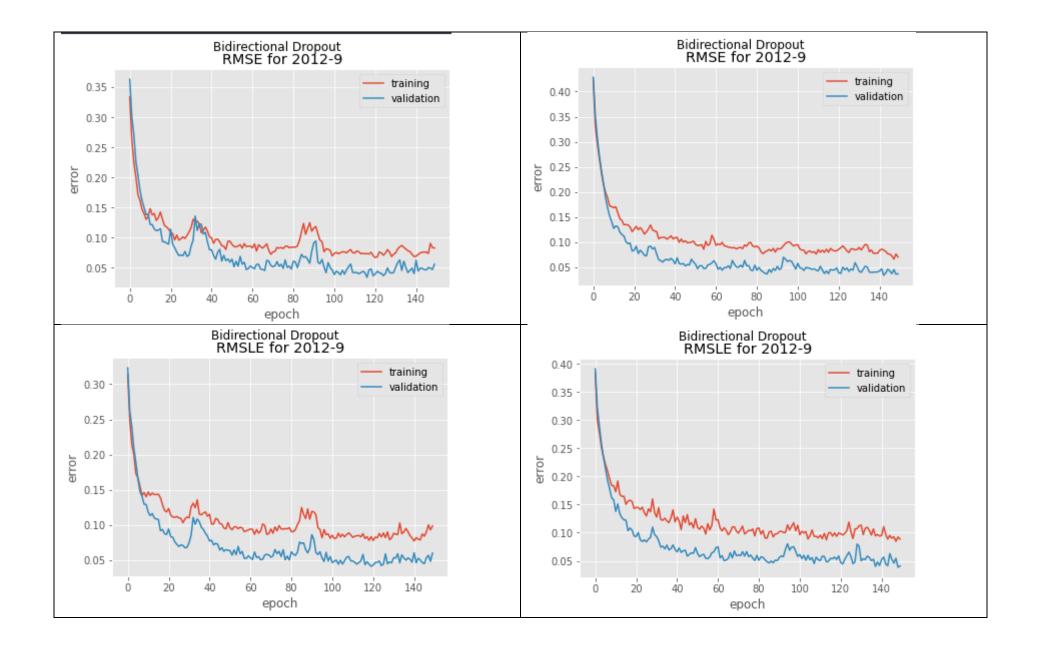


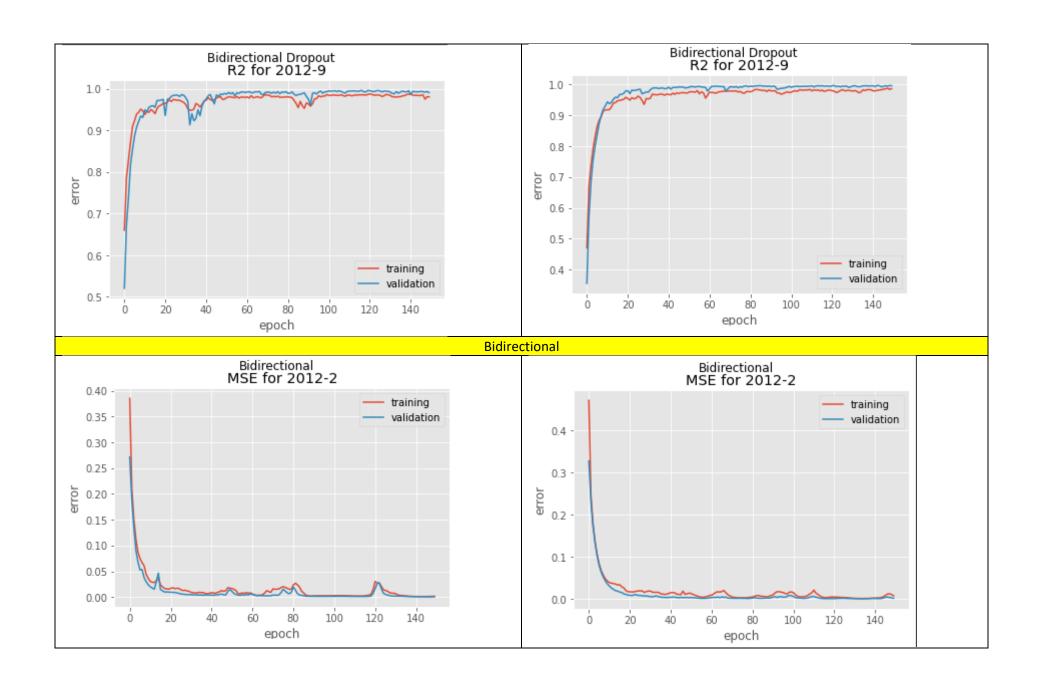


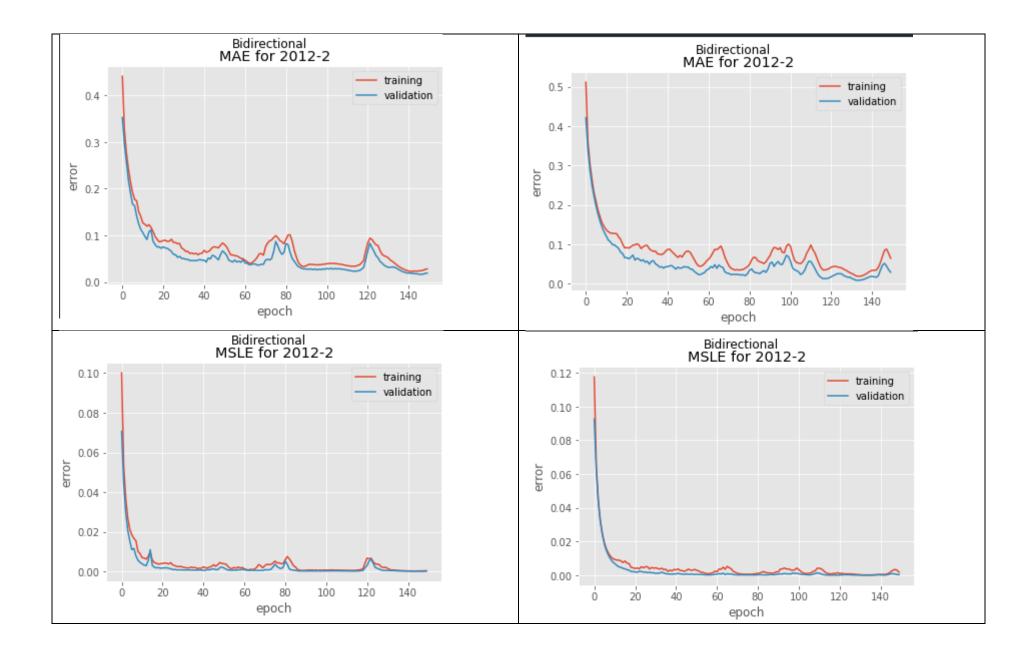


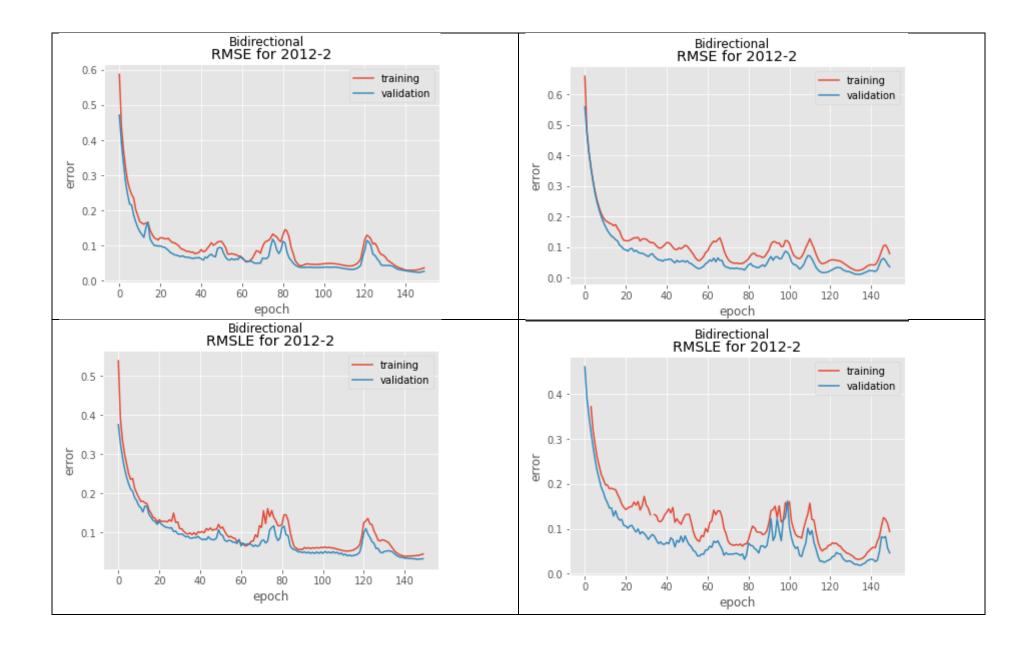


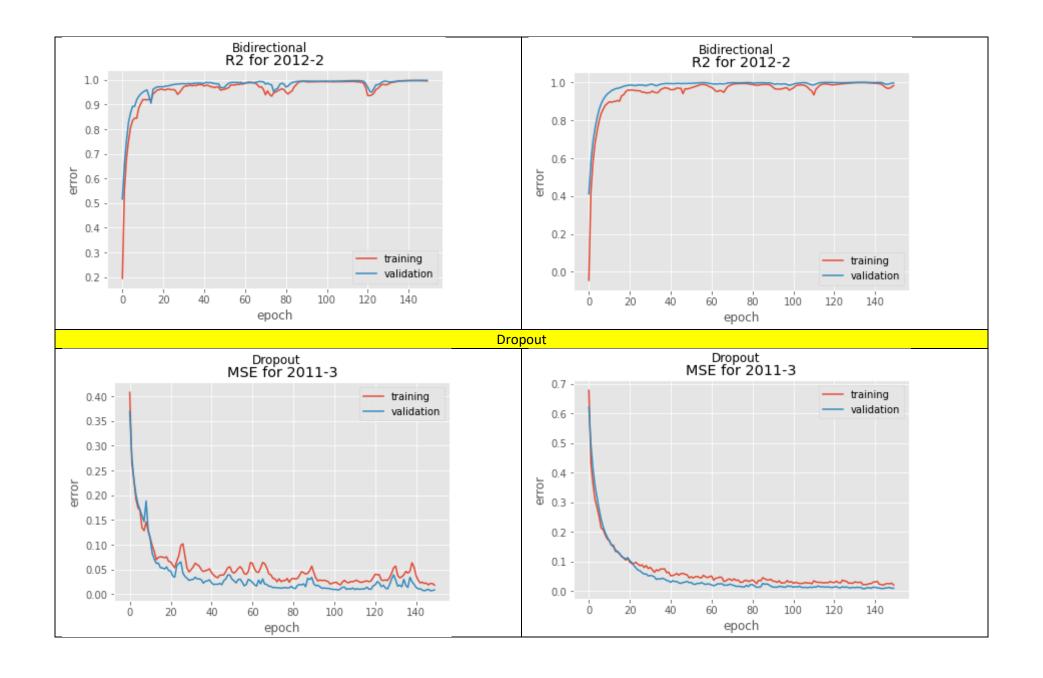


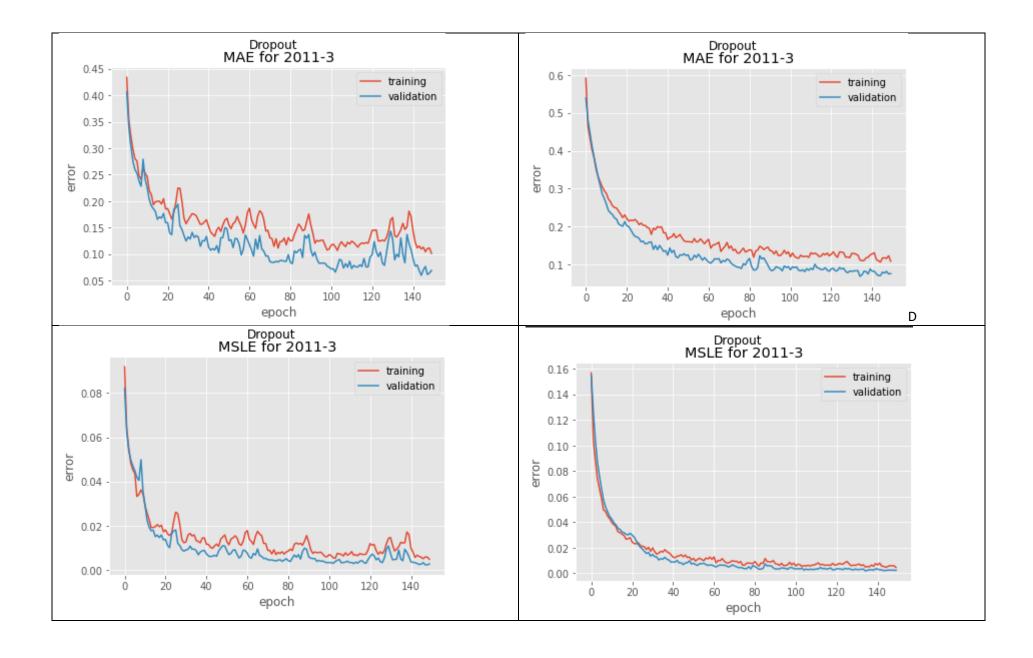


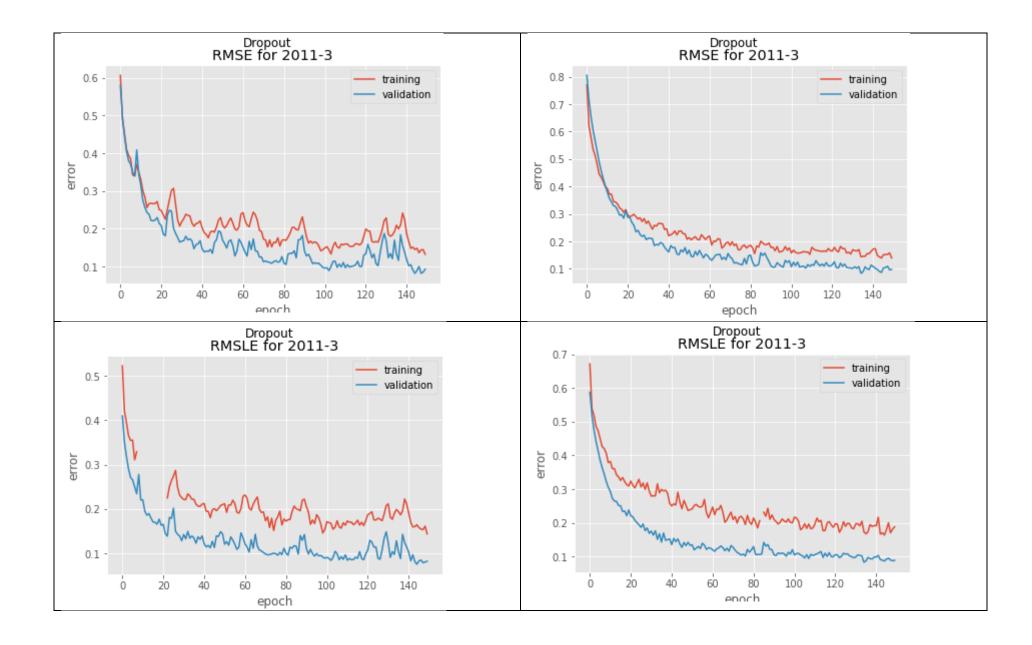


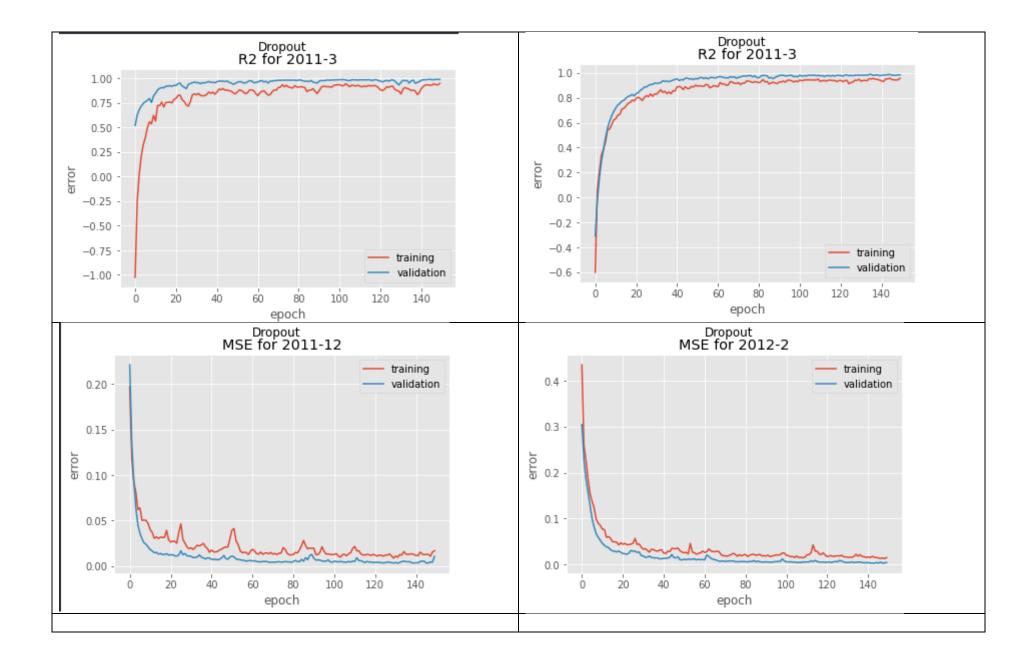


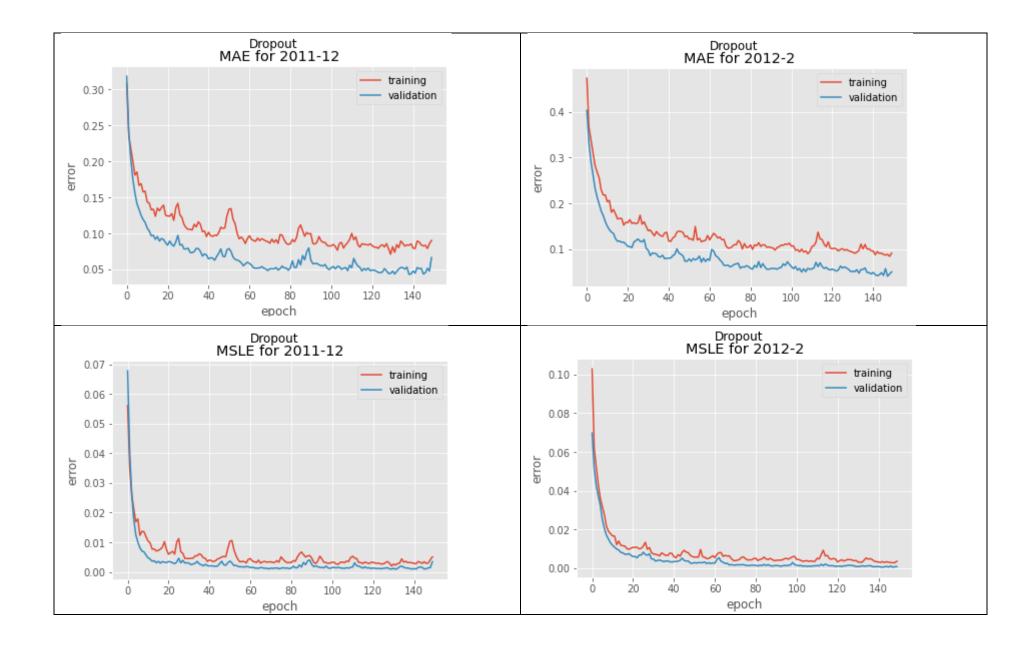


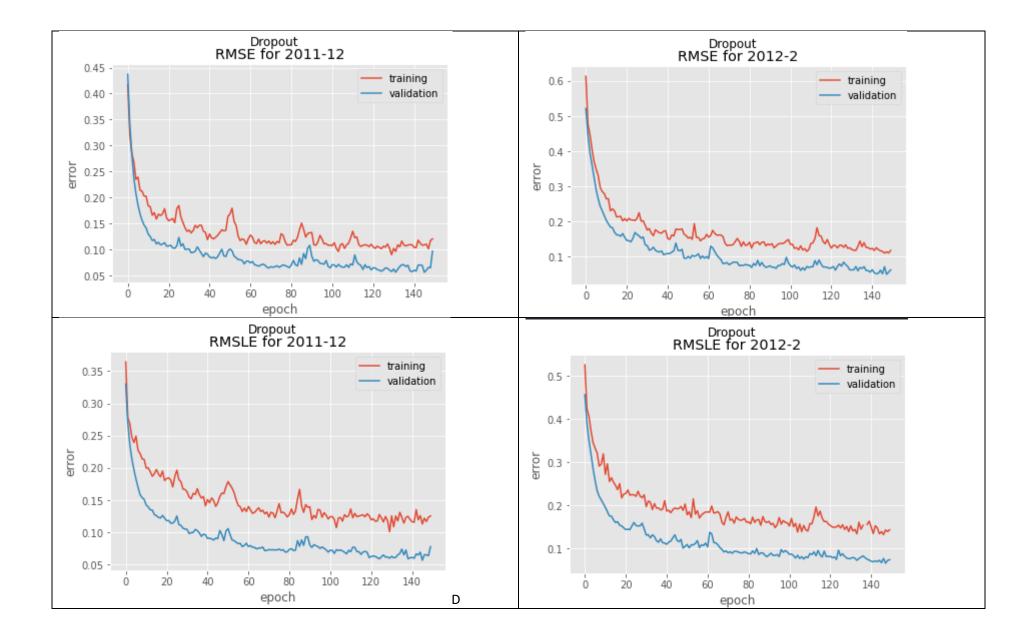


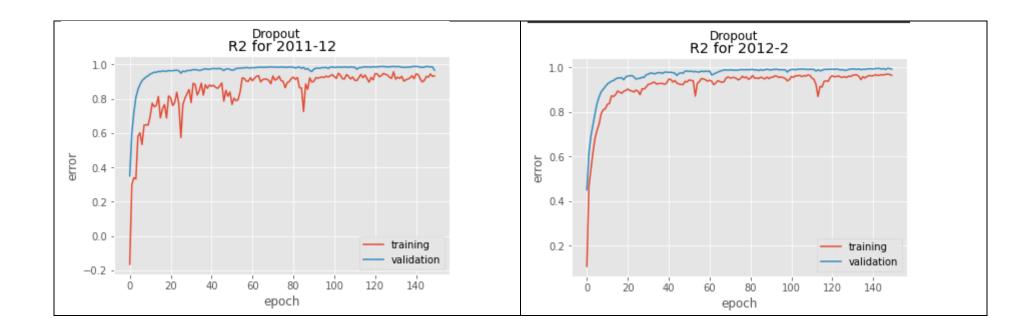


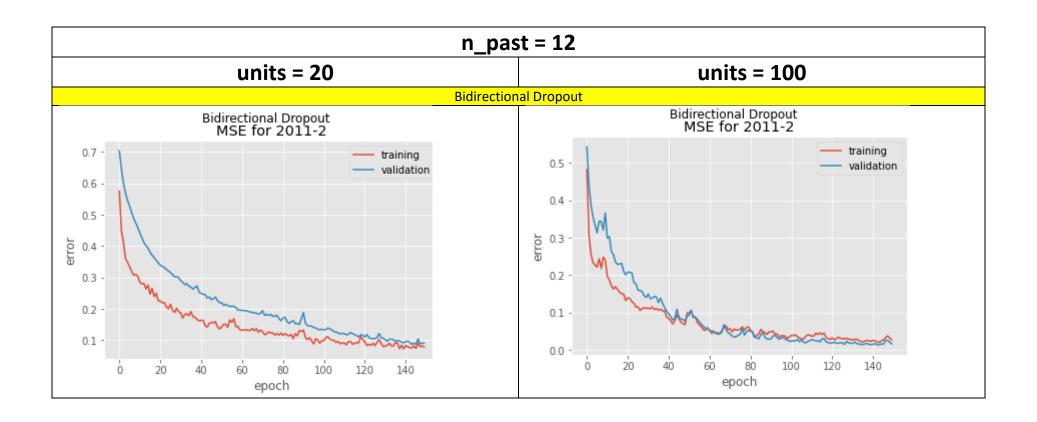


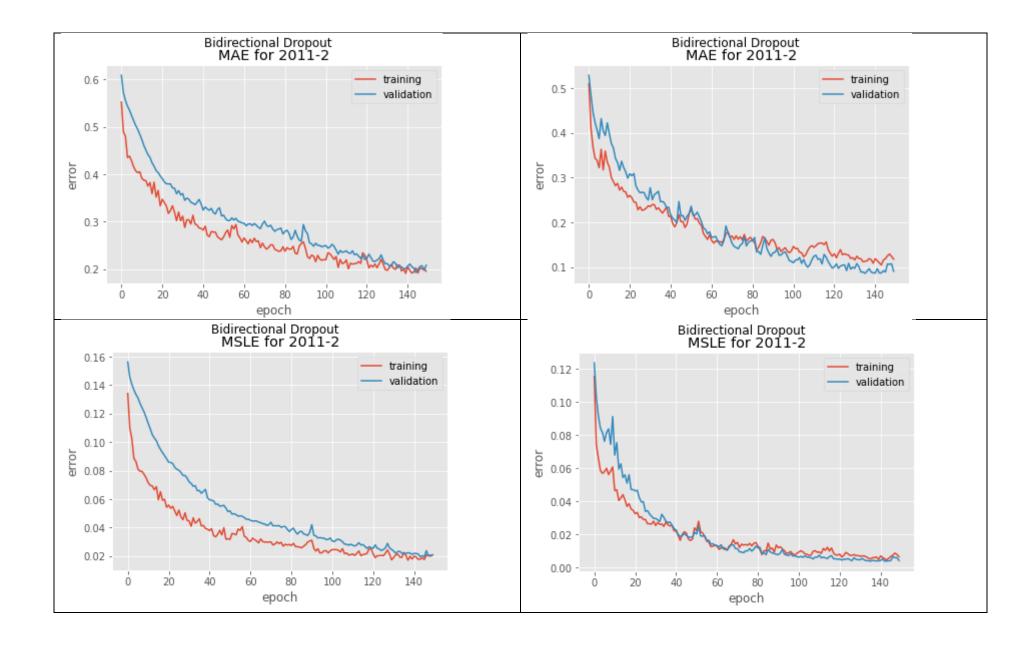


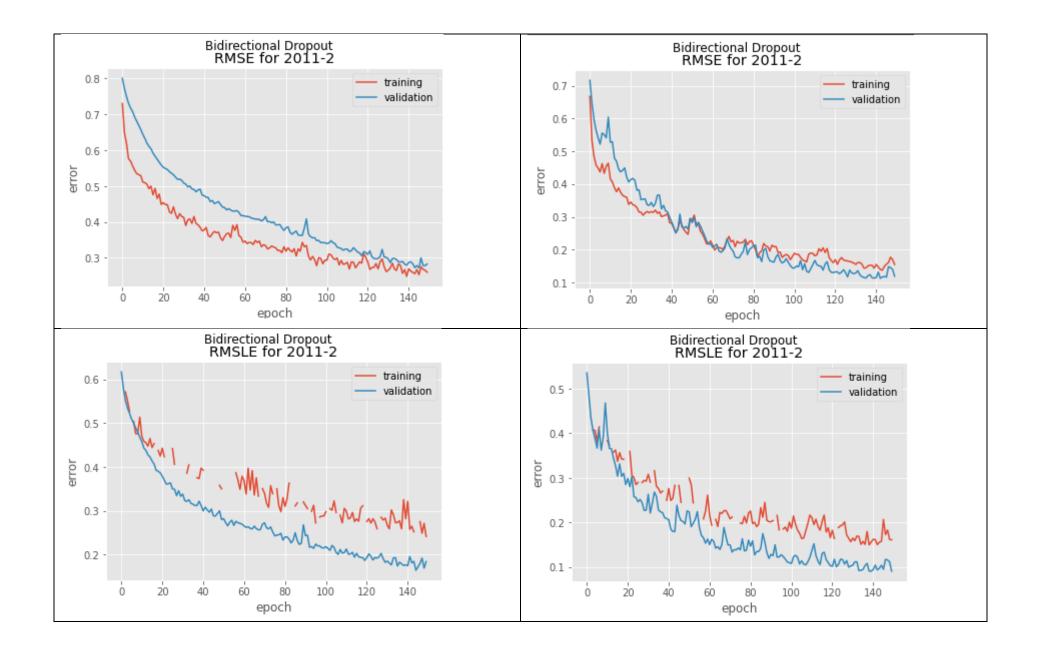


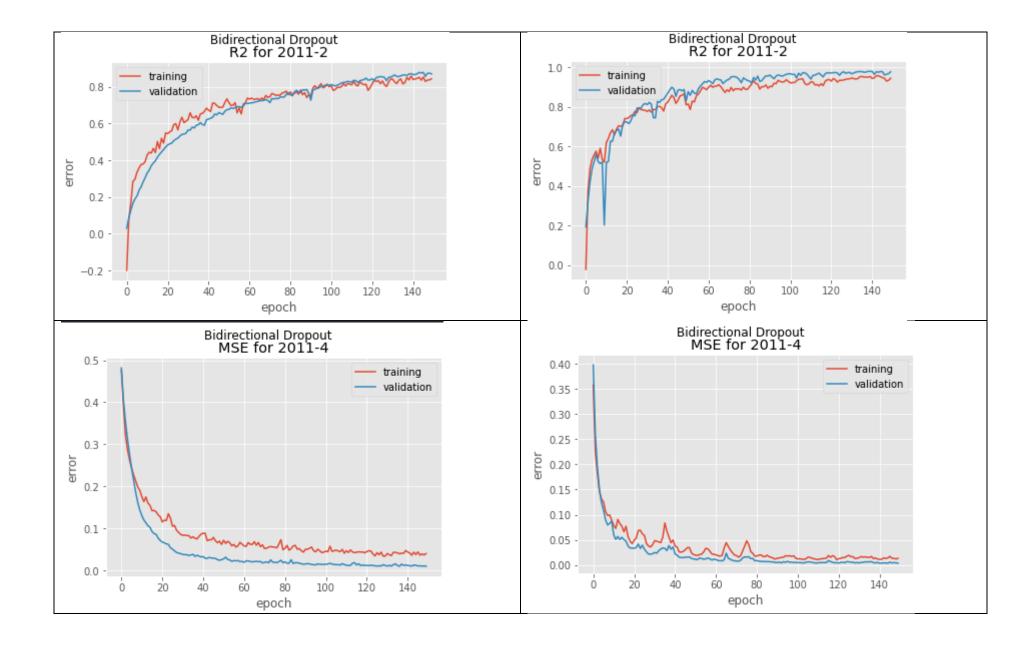


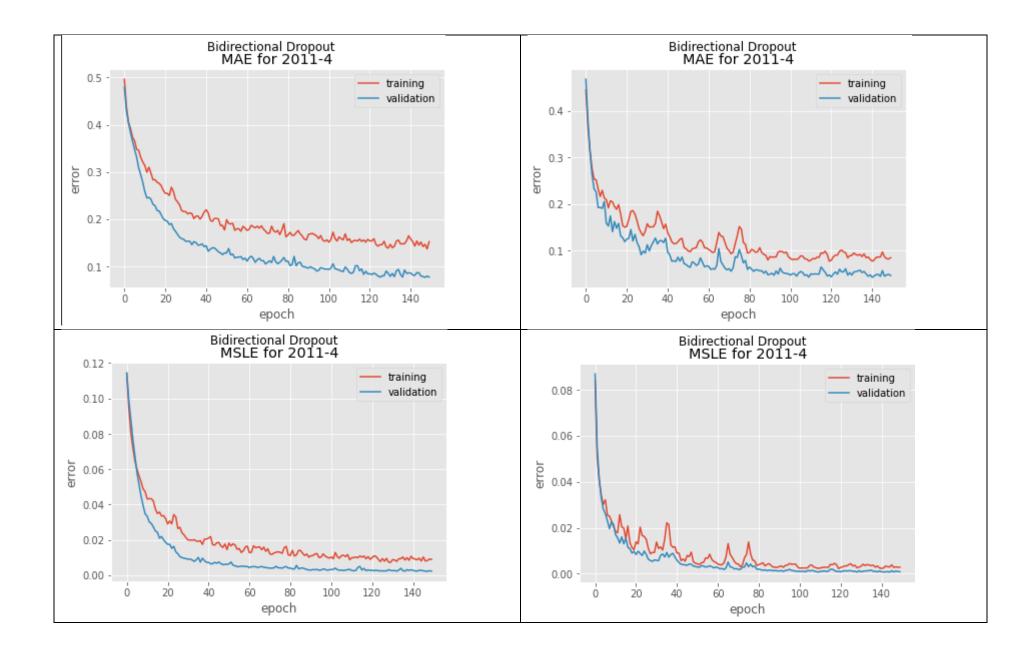


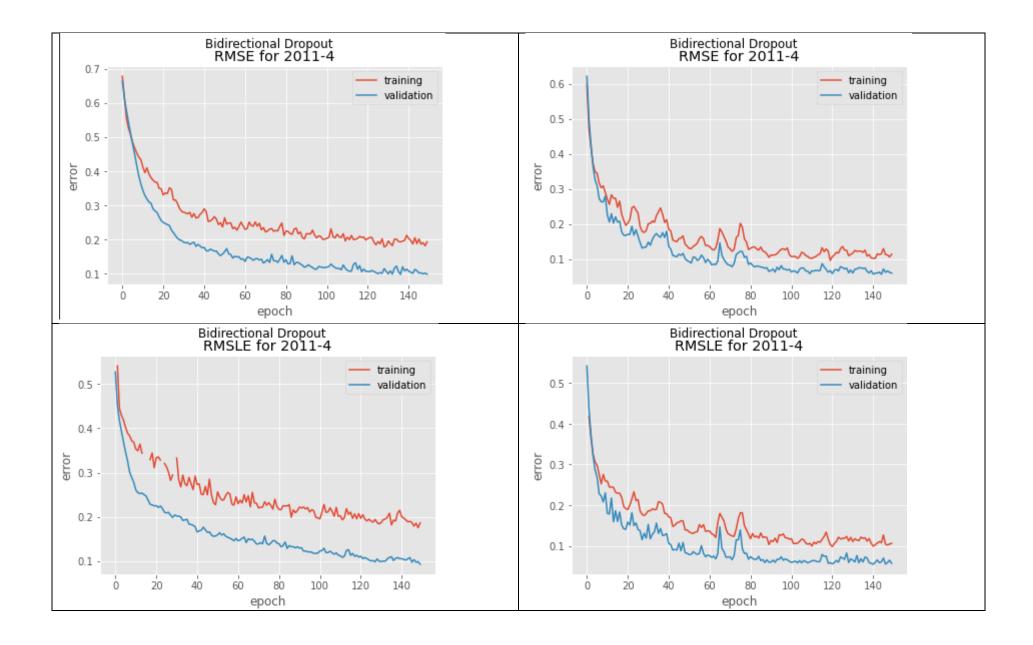


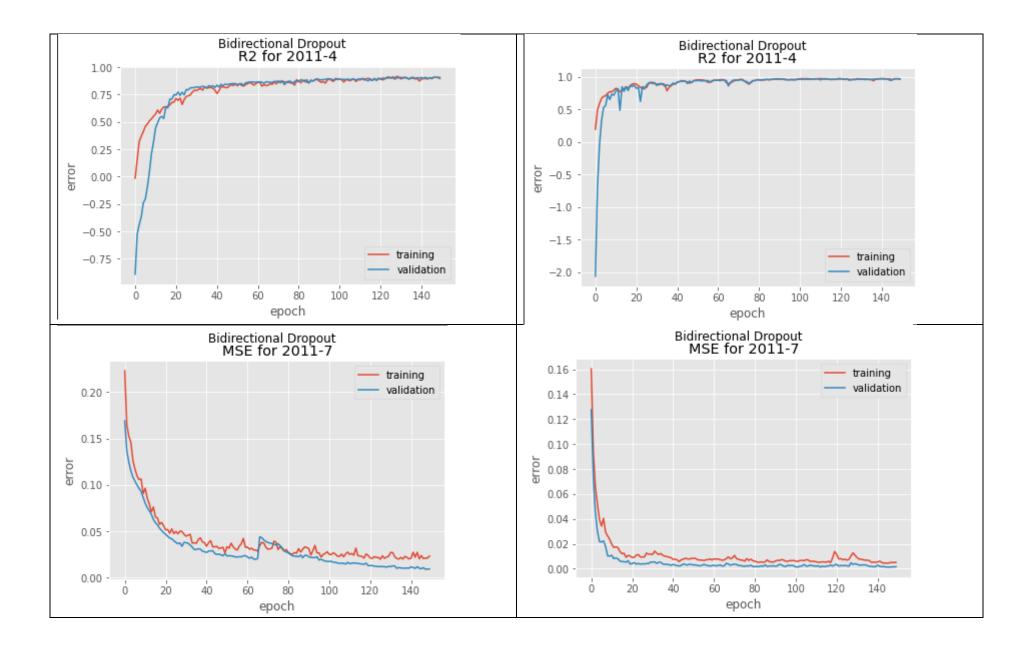


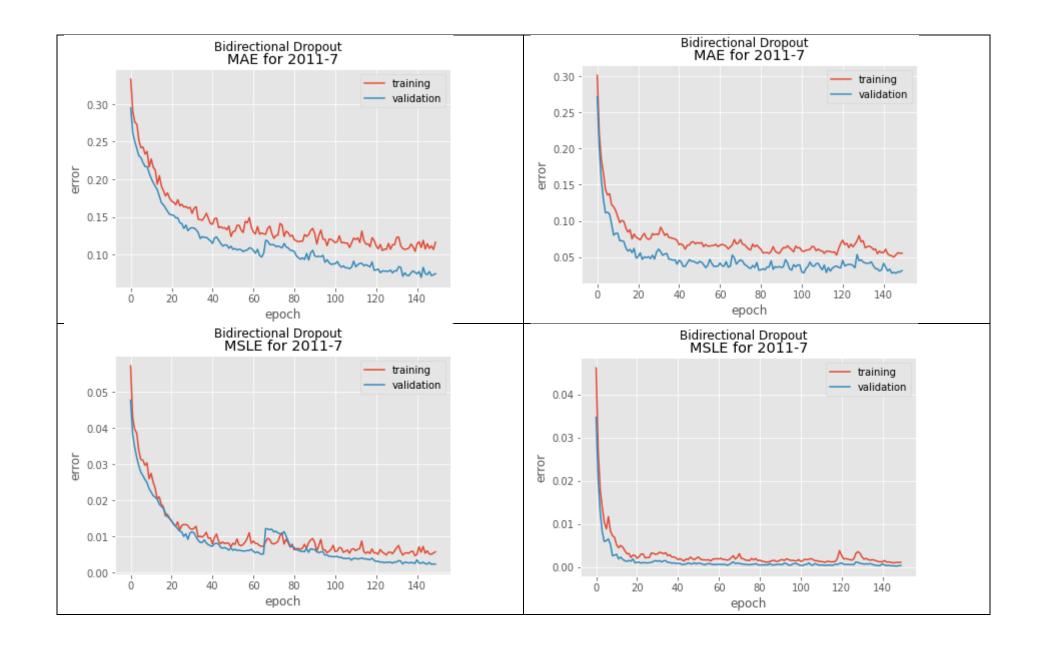


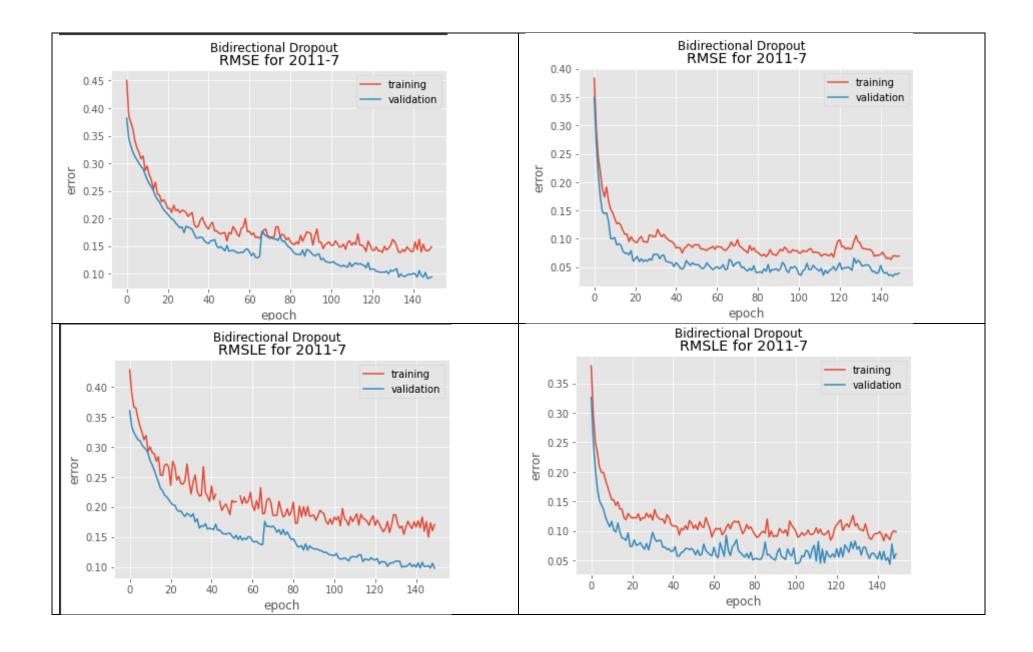


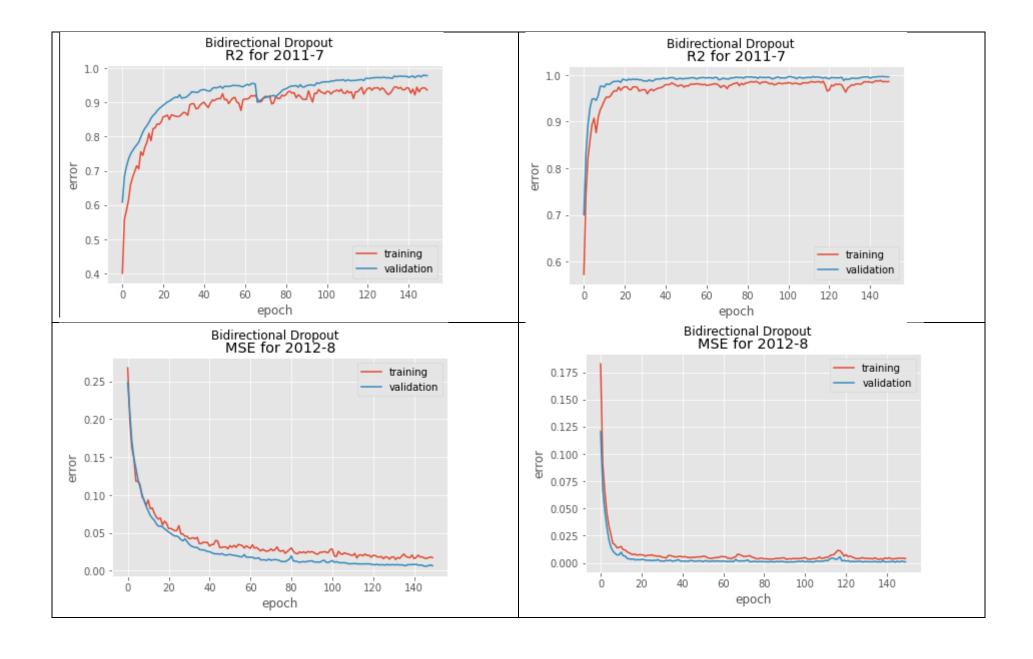


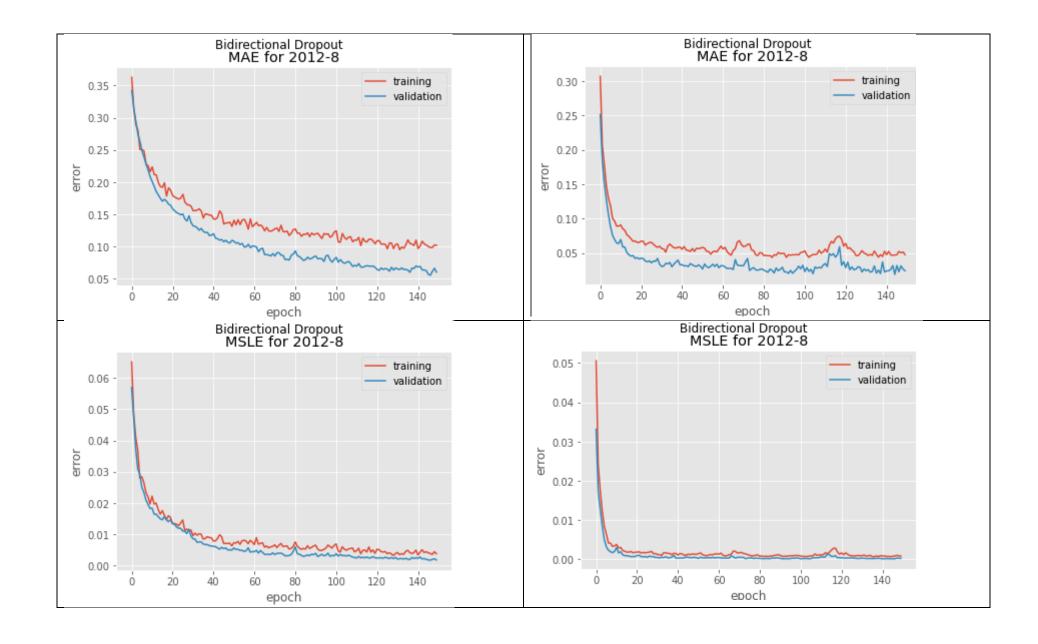


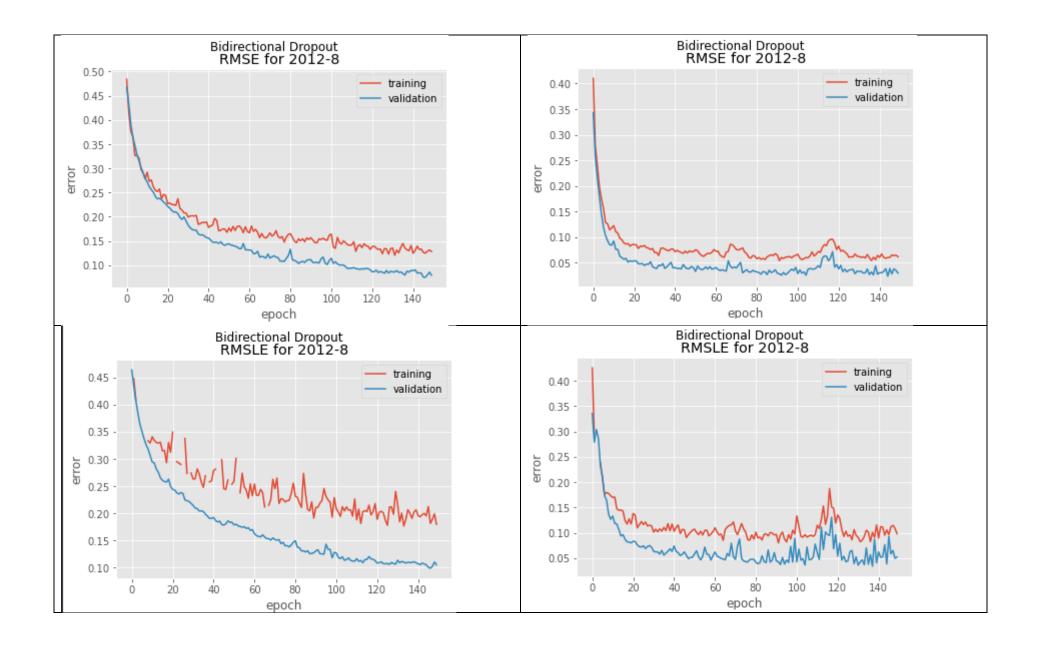


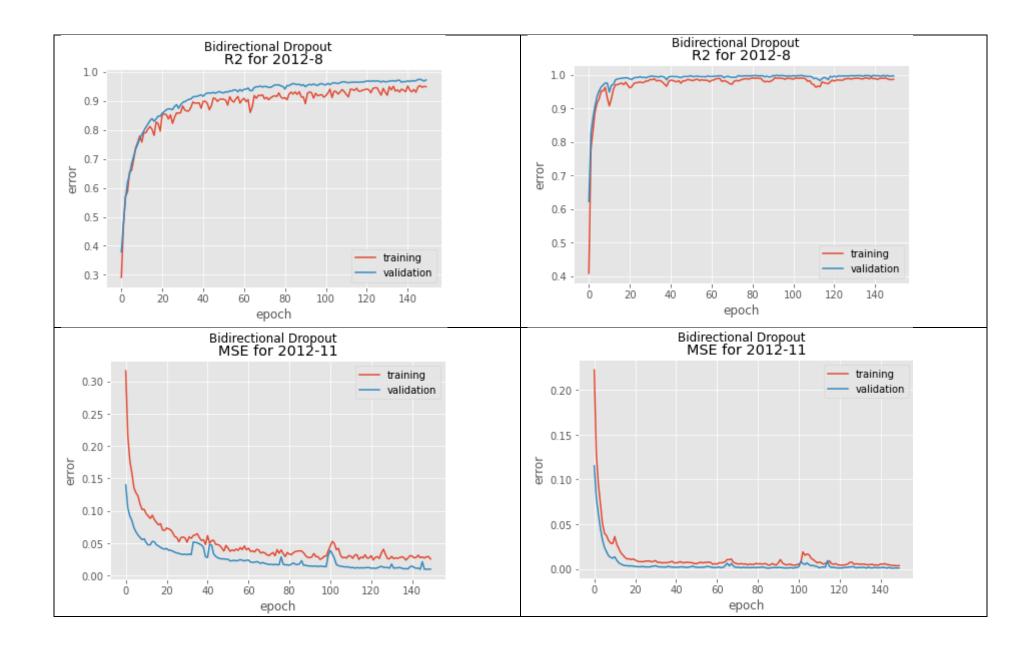


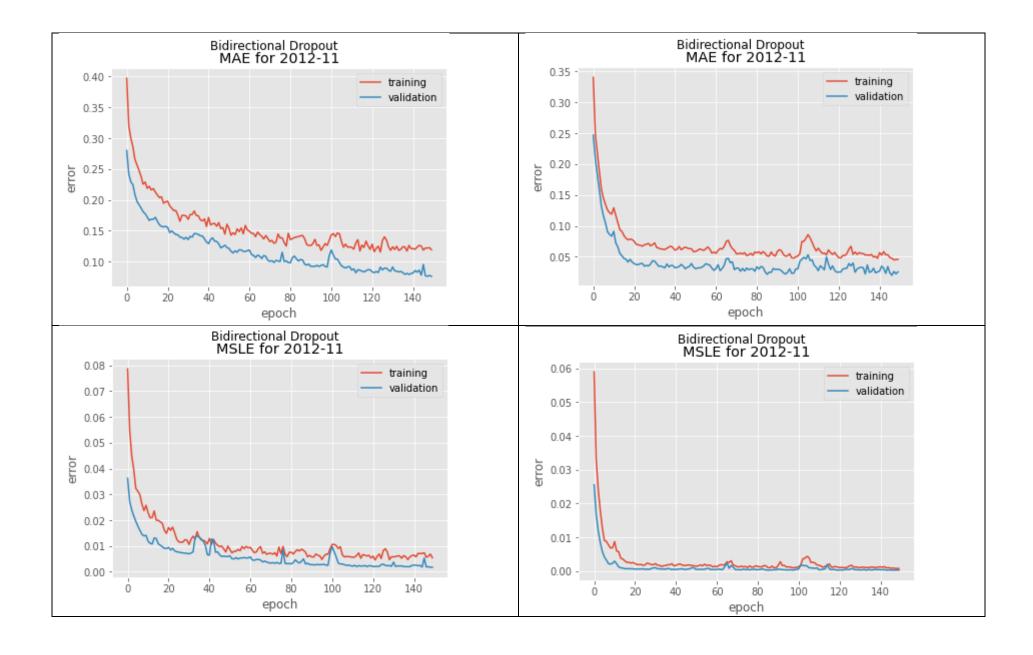


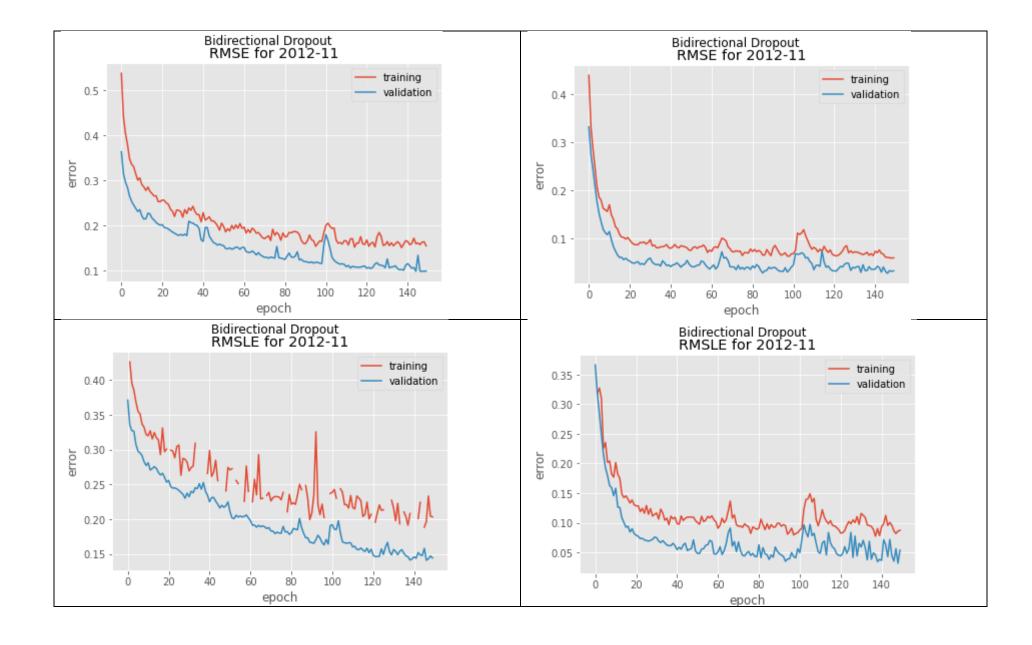


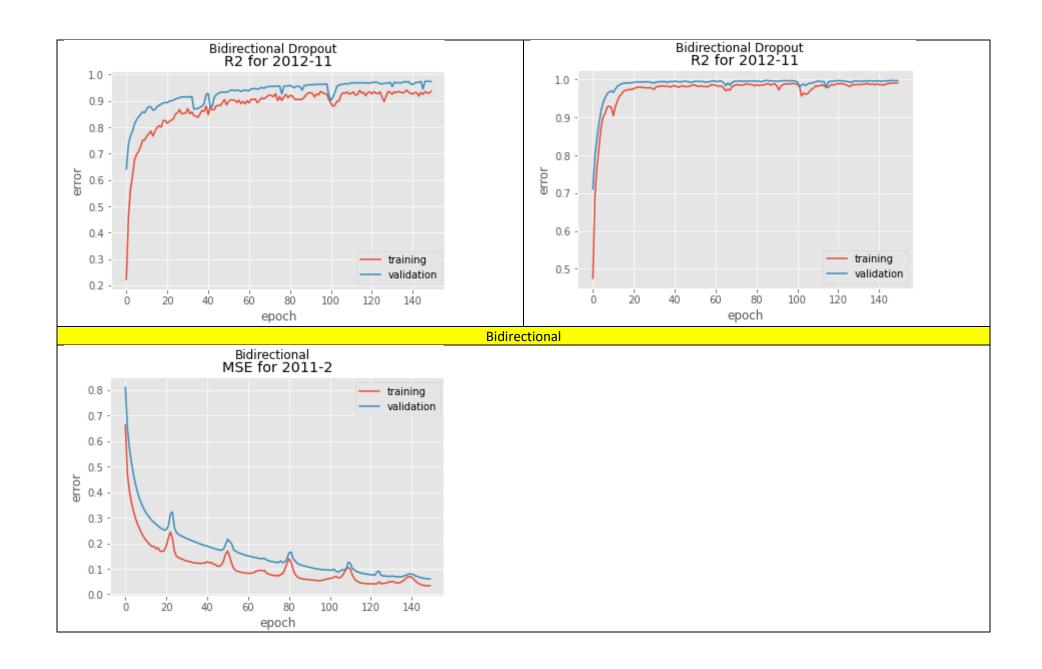


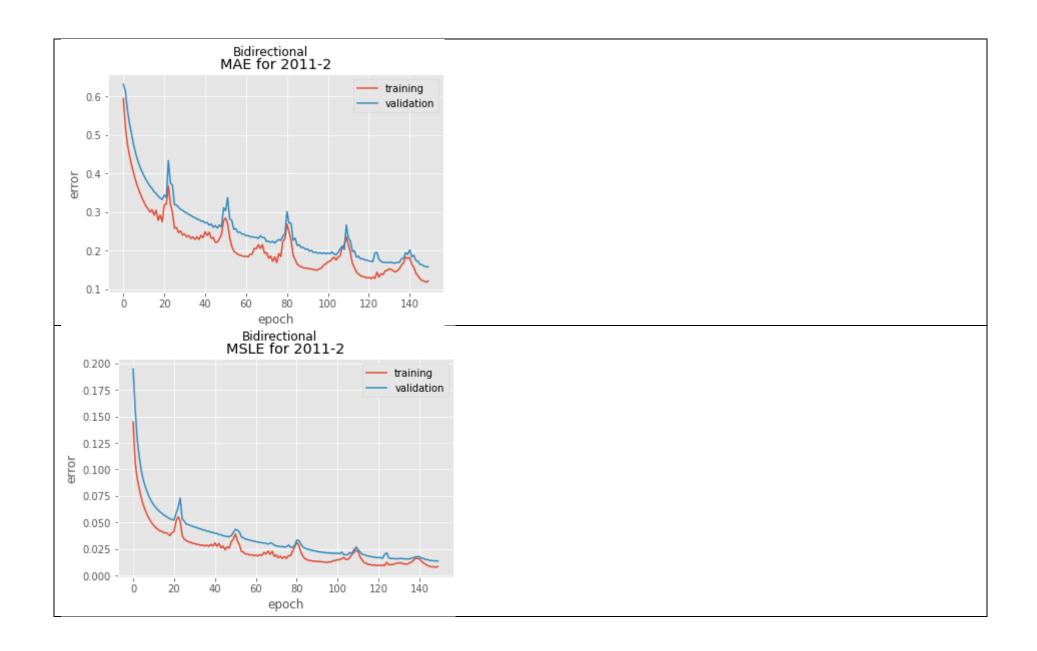


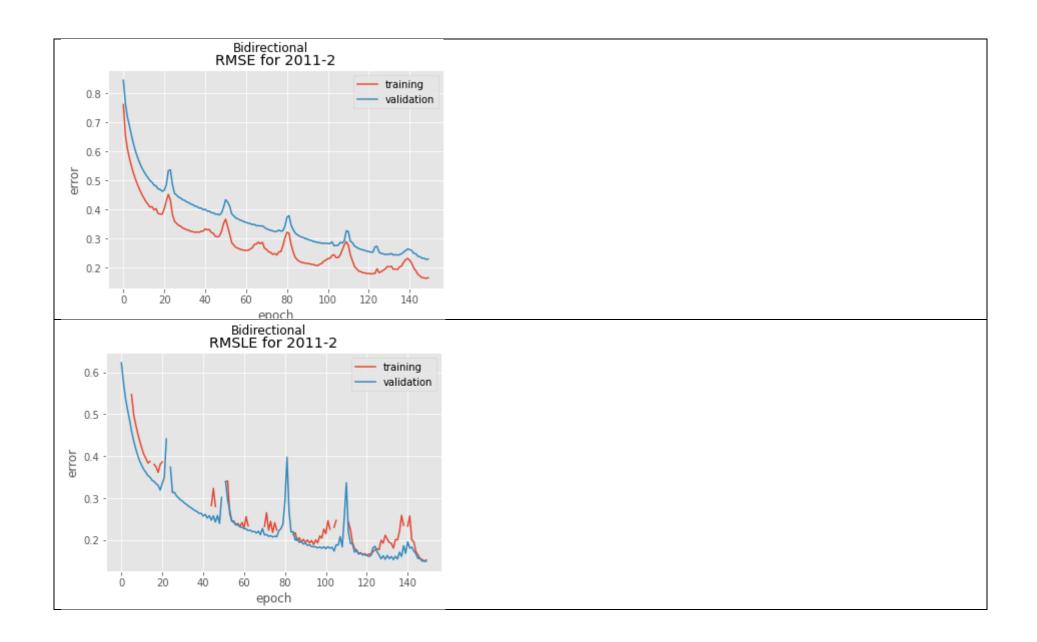


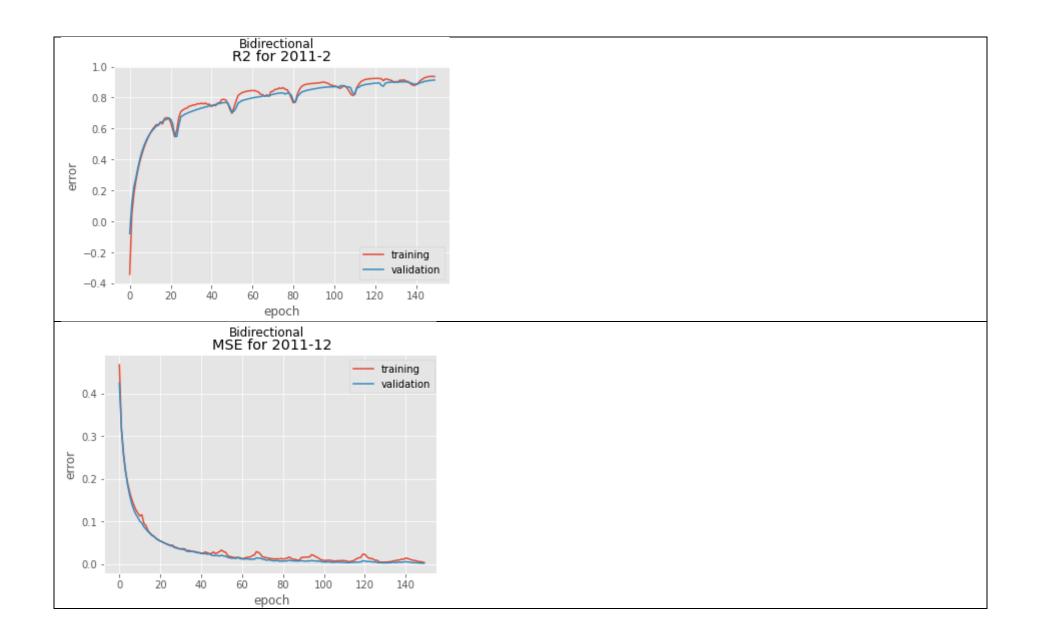


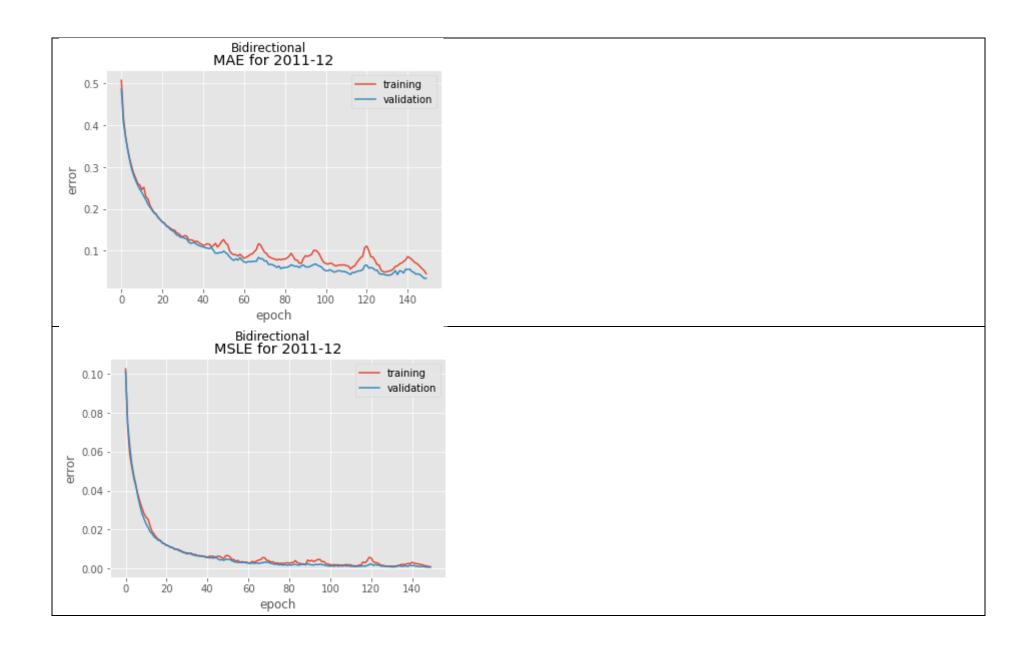


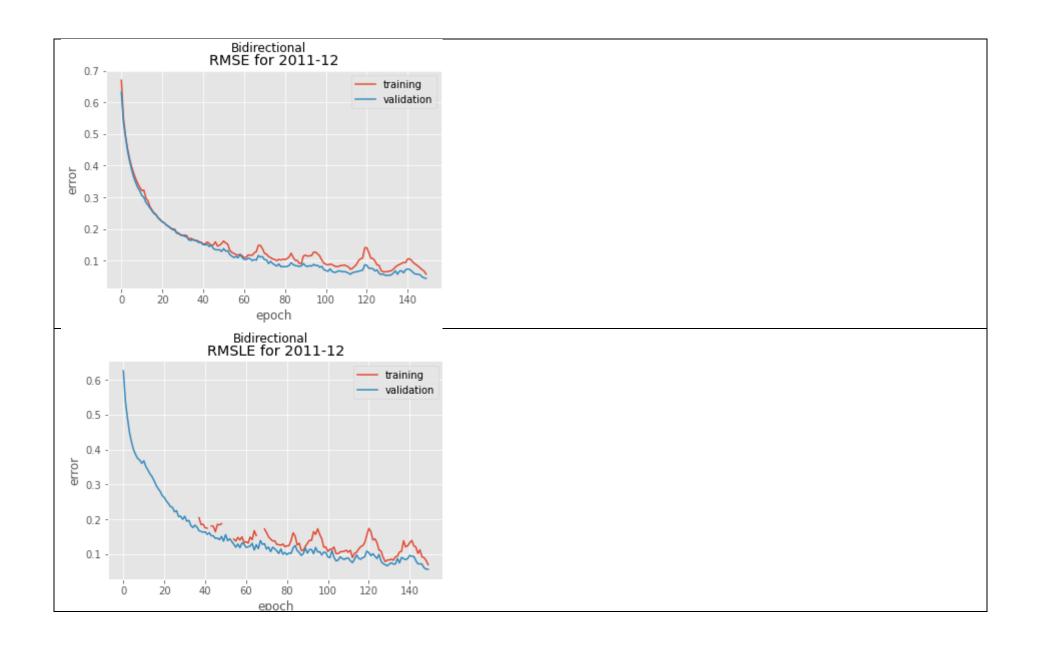


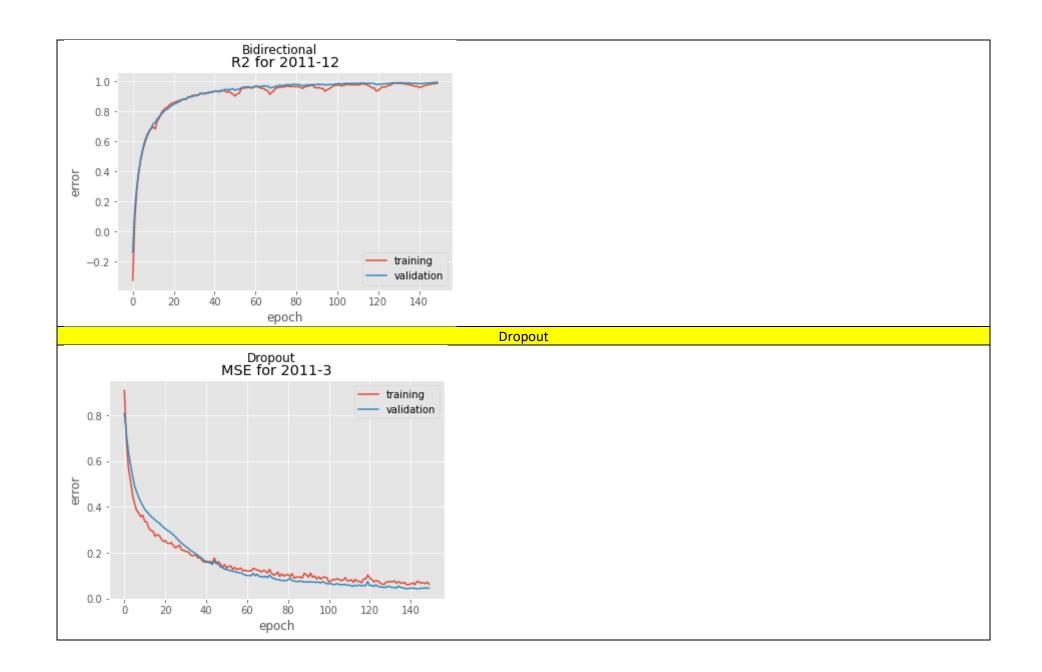


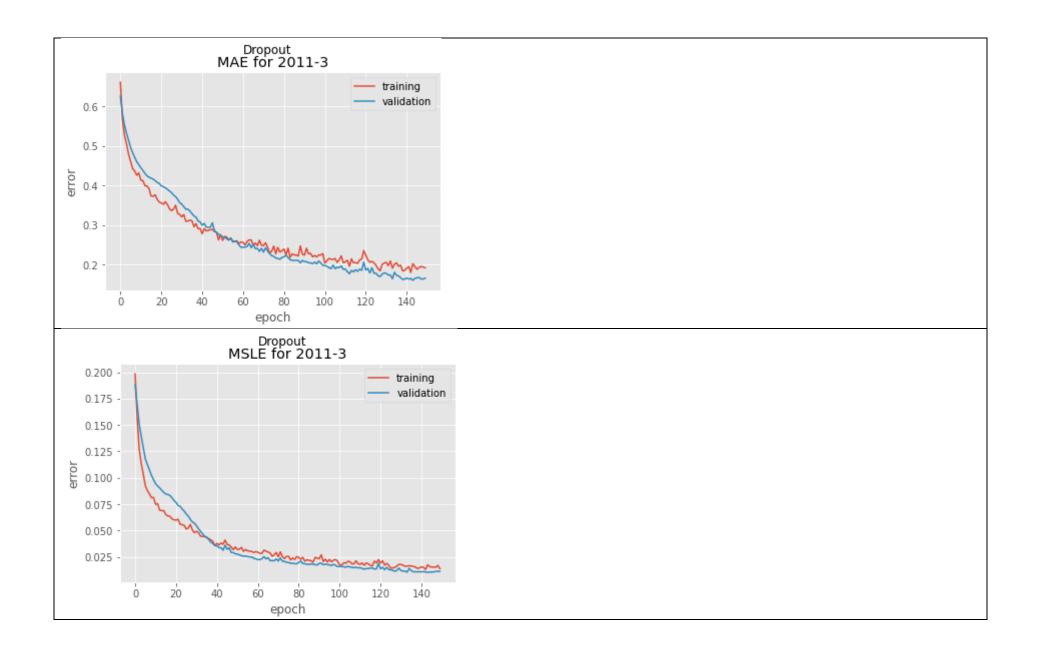


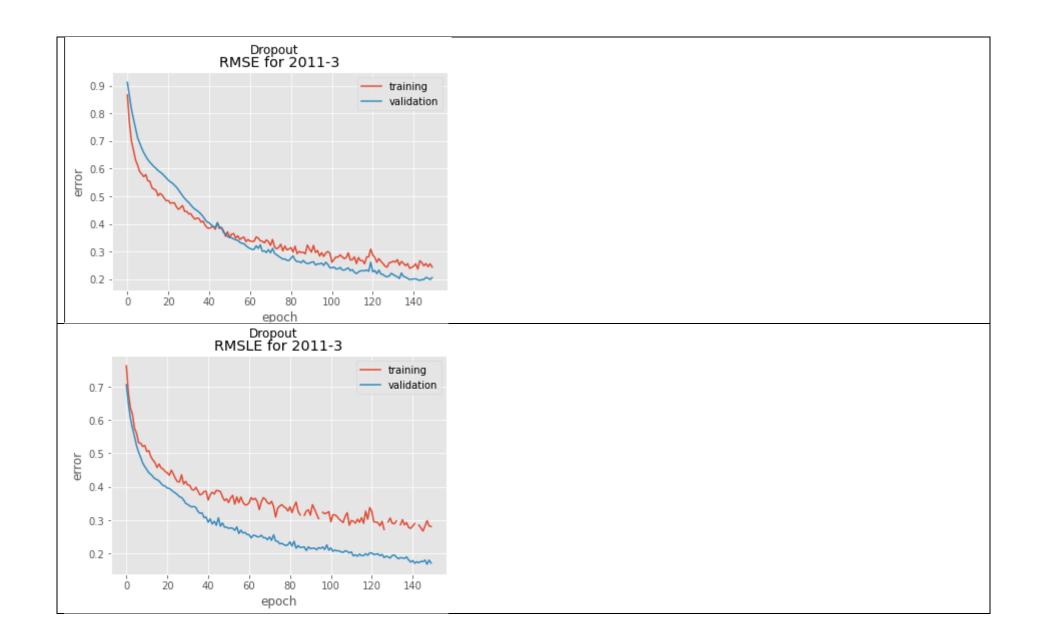


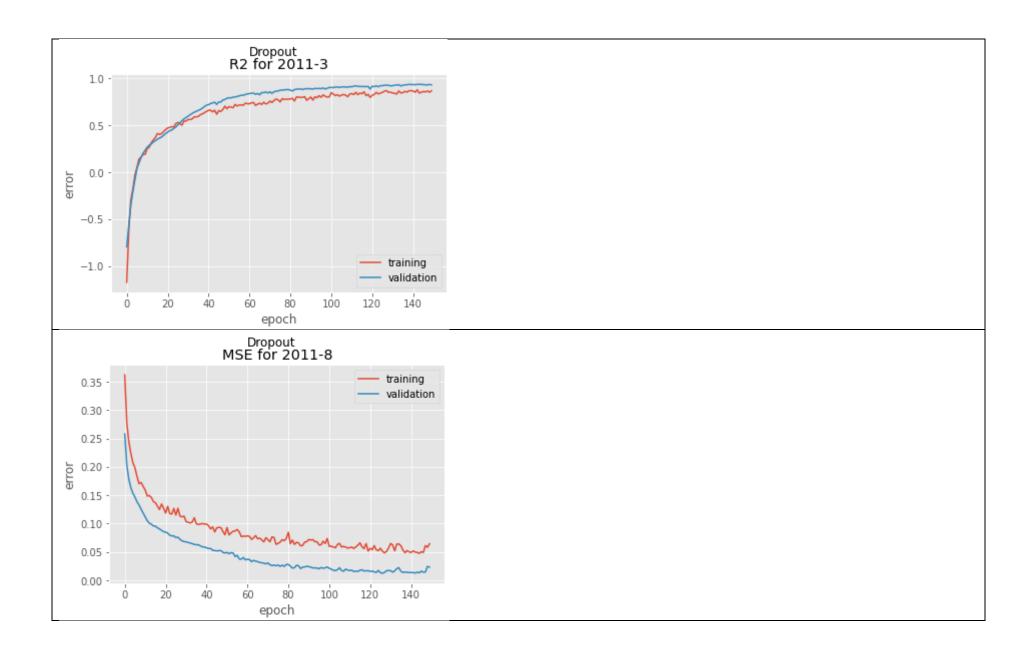


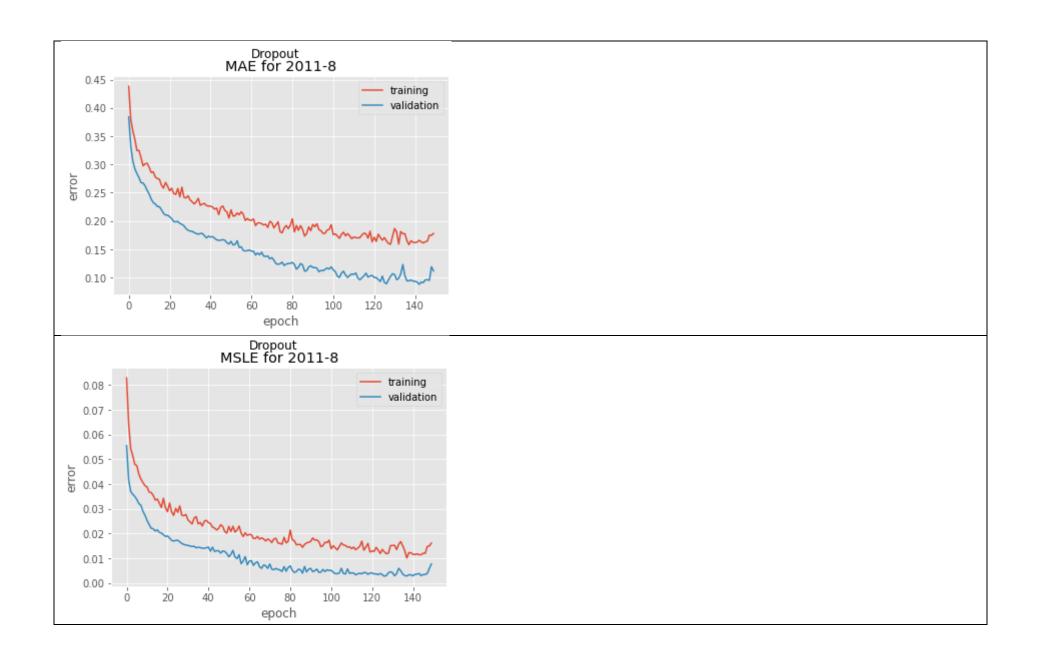


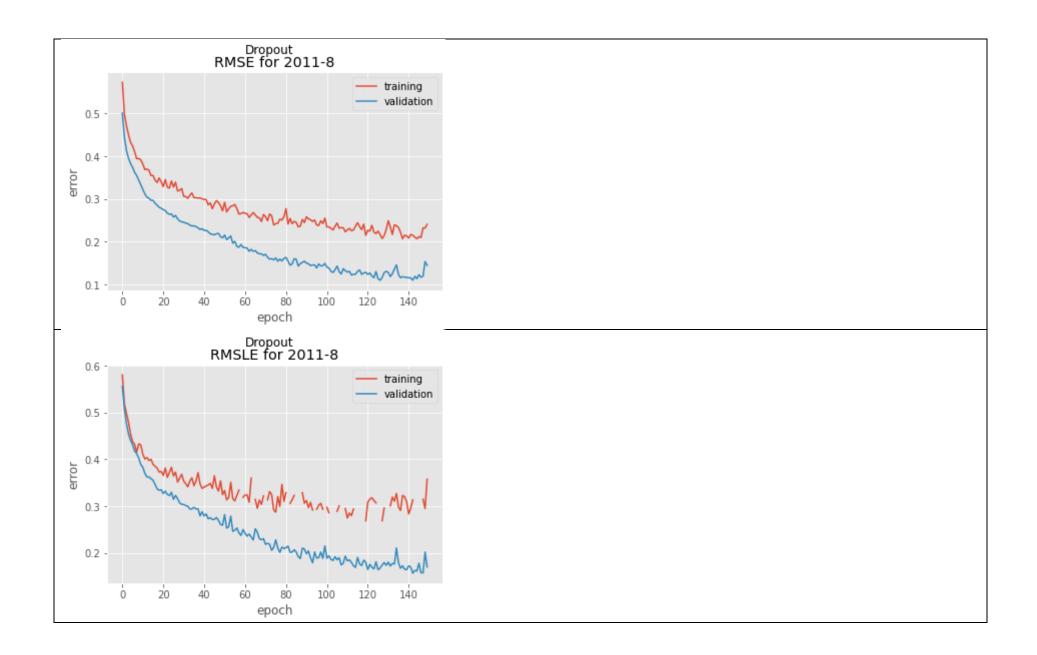


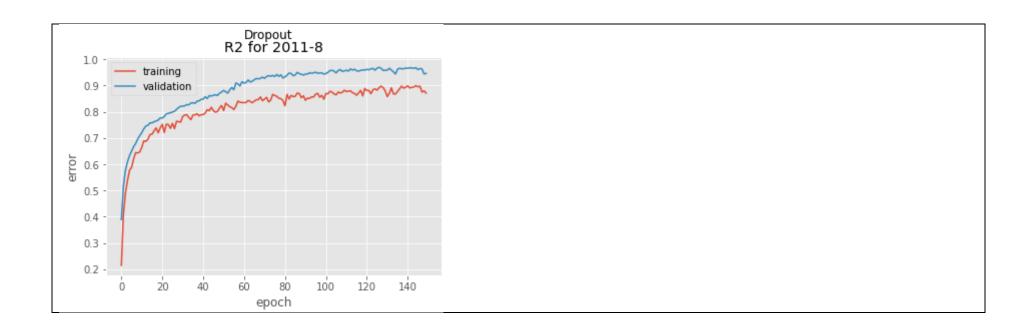












LATE SUBMISSIONS

n_past = 24 UNITS = 50

Submission and Description	Private Score
submission_Bidirectional.csv	1.53835
submission_Bidirectional Dropout.csv	1.49210

n_past = 12 UNITS = 50

submission_Bidirectional.csv	1.56224
submission_Bidirectional Dropout.csv	1.48797
submission_Dropout.csv	1.42280

n_past = 12 UNITS = 10

submission_Bidirectional.csv 1.58528

submission_Bidirectional Dropout.csv

1.40243

submission_Dropout.csv

1.37222

n past = 12 UNITS = 100

submission_Bidirectional Dropout.csv

1.54456

CONCLUSIONS

If n_past = 24 => more jagged curves, many "bumps" and "valleys"

If n_past = 12 => less jagged curves, fewer "bumps" and "valleys"

Models without "Bidirectional" and with "Dropout" seemed to be less jagged, and seemed to perform better.

Models with less units seemed to perform better.

Models with higher number of units took way longer to train.

Gaps in RMSLE plot are caused by linear activation function (default in keras if not specified), which can have positive and negative values (we can log only positive values). A much better alternative is RELU activation function.

Use MinMaxScaler instead of RoboustScaler. RoboustScaler caused forecast plot to have a minimal value other than 0 (like 100, 150, 220). This make the plot look bad and gave inaccurate predictions.

Conclusions from notebook 003a: train_X dataset was used to train the models instead of trainX. This was the cause of no curve convergence.