1 Tuned parameters

Table 1: Candidate models for NH_3N forecasting[1].

Machine learning techniques	Model names
Ensemble learning	RF
Deep learning	CNN
Deep learning	DNN
Deep learning	RNN
Deep learning	GRU
Deep learning	LSTM

[1] All the hyper-parameters are set to be the same. Hidden layer = 10, output layer = 3, learning rate = 5e-05, epoch = 100, experiment times = 3, and scheduler is used for regularization (patience=10, factor=0.5).

Table 2: Configurations for NH_3N forecasting.

Parameter	Values	Pre-processing methods	Parameters
Input	24 hours	SG 1	span=5
Output	3 hours	SG 2	span=7
Train	12/23/21 to 1/09/22	SG 3	span=9
Valid	1/10/22 to 1/15/22	EWMA 1	span=2
Test	1/16/22 to 1/22/22	EWMA 2	span=3
		EWMA 3	span=4
		OR[2]	-

[2]Three days were identified as outliers and removed from the training dataset.

2 Reuslt 1 (Exp-1)

2.1 test and val loss comparison in Jan

- First time showing the results in decending order of test loss.
- LSTM and GRU have lower test loss than RNN, DNN, and RF.
- The lowest test loss of NH_3N forecasting approach has higher validation loss than several approaches.

Table 3: Evaluation of each baseline approach for NH_3N forecasting.

Rank	Model-Dataset	Test loss*	valid loss
1	GRU-sg7	0.0383 ±0.0007	1.2508 ±0.0458
2	GRU-sg5	0.0385 ±0.0001	1.2644 ±0.0081
3	LSTM-ew3	0.0388 ±0.0006	**1.0796 ±0.0112**(1)
4	LSTM-sg7	0.0388 ±0.0003	1.1804 ±0.0296
5	LSTM-sg5	0.0388 ±0.0003	1.2346 ±0.0520
6	GRU-ew2	0.0389 ±0.0004	1.1891 ±0.0307
7	GRU-ew4	0.0391 ±0.0004	1.2390 ±0.0557
8	LSTM-ew2	0.0392 ±0.0006	**1.0969 ±0.0159**(2)
9	GRU-ew3	0.0392 ±0.0002	1.2199 ±0.0137
10	LSTM-ew4	0.0395 ±0.0010	**1.1219 ±0.0079**(3)
11	GRU-sg9	0.0396 ±0.0003	1.3097 ±0.0175
12	LSTM-or	0.0398 ±0.0003	1.2612 ±0.0269
13	LSTM-obs	0.0405 ±0.0004	1.2366 ±0.0150
14	GRU-or	0.0405 ±0.0002	1.3993 ±0.0532
15	LSTM-sg9	0.0410 ±0.0005	1.3076 ±0.0214
16	GRU-obs	0.0414 ±0.0005	1.3638 ±0.0359
17	RNN-sg5	0.0415 ±0.0001	1.5088 ±0.0336
18	RNN-ew2	0.0421 ±0.0007	1.5425 ±0.0566
19	RNN-sg7	0.0423 ±0.0008	1.6267 ±0.0065
20	RNN-ew4	0.0432 ±0.0003	1.5992 ±0.0300

2.2 Test loss in Oct and val

- Result 1 (baseline performance in October)

Table 4: Model evaluations of each forecasting approach.

Rank	Model-Dataset	Test loss*	valid loss
1	LSTM-ew3	0.0158	1.0796
2	LSTM-ew2	0.0161	1.0969
3	LSTM-ew4	0.0163	1.1219
4	LSTM-sg5	0.0166	1.2346
5	GRU-ew3	0.0167	1.2199
6	GRU-ew4	0.0169	1.2390
7	GRU-ew2	0.0170	1.1891
8	GRU-sg9	0.0174	1.3097
9	LSTM-obs	0.0175	1.2366
10	LSTM-or	0.0177	1.2612
11	GRU-sg5	0.0178	1.2644
12	GRU-sg7	0.0180	1.2508
13	LSTM-sg7	0.0180	1.1804
14	GRU-or	0.0187	1.3993
15	LSTM-sg9	0.0188	1.3076
16	GRU-obs	0.0189	1.3638
17	RNN-ew4	0.0190	1.5992
18	RNN-ew2	0.0191	1.5425
19	RNN-ew3	0.0193	1.6041
20	RNN-sg5	0.0195	1.5088

2.3 Point out the top forecasting approach of the test loss didn't have the lowest validation loss.

- Test dataset from 1/16 to 1/22 performed differently on the same forecasting approach compared to validation loss.

Table 5: Comparison of NH_3N val/test loss from 1/16 to 1/22.

GRU	Test loss*	Val loss	LSTM	Test loss*	Val loss
sg7	0.0383±0.0007	1.2508±0.0458	ew3	0.0388±0.0006	**1.0796±0.0112**(1)
sg5	0.0385±0.0001	1.2644±0.0081	sg7	0.0388±0.0003	1.1804±0.0296
ew2	0.0389±0.0004	**1.1891±0.0307**(1)	sg5	0.0388±0.0003	1.2346±0.0520
ew4	0.0391±0.0004	**1.2390±0.0557**(3)	ew2	0.0392±0.0006	**1.0969±0.0159**(2)
ew3	0.0392±0.0002	**1.2199±0.0137**(2)	ew4	0.0395±0.0010	**1.1219±0.0079**(3)
sg9	0.0396±0.0003	1.3097±0.0175	or	0.0398±0.0003	1.2612±0.0269
or	0.0405±0.0002	1.3993±0.0532	obs	0.0405±0.0004	1.2366±0.0150
obs	0.0414±0.0005	1.3638±0.0359	sg9	0.0410±0.0005	1.3076±0.0214

2.4 Test dataset from 10/10 to 10/16 performed similar on the same forecasting approach compared to validation loss.

Table 6: Val/test loss of NH_3N from 10/10 to 10/16.

GRU	Test loss	Val loss	LSTM	Test loss	Val loss
ew3	0.0167±0.0000	**1.2199±0.0137**(2)	ew3	0.0158±0.0004	**1.0796±0.0112**(1)
ew4	0.0169±0.0001	**1.2390±0.0557**(3)	ew2	0.0161±0.0000	**1.0969±0.0159**(2)
ew2	0.0170±0.0004	**1.1891±0.0307**(1)	ew4	0.0163±0.0003	**1.1219±0.0079**(3)
sg9	0.0174±0.0002	1.3097±0.0175	sg5	0.0166±0.0001	1.2346±0.0520
sg5	0.0178±0.0004	1.2644±0.0081	obs	0.0175±0.0001	1.2366±0.0150
sg7	0.0180±0.0005	1.2508±0.0458	or	0.0177±0.0002	1.2612±0.0269
or	0.0187±0.0002	1.3993±0.0532	sg7	0.0180±0.0002	1.1804±0.0296
obs	0.0189±0.0002	1.3638±0.0359	sg9	0.0188±0.0002	1.3076±0.0214

2.5 The influence of each pre-processing method on model training is different.

Table 7: Evaluation of pre-processing methods on LSTM and GRU.

Rank	GRU[3]	LSTM[3]	GRU[4]	LSTM[4]
1	sg7	ew3	ew3	ew3
2	sg5	sg7	ew4	ew2
3	ew2	sg5	ew2	ew4
4	ew4	ew2	sg9	sg5
5	ew3	ew4	sg5	obs
6	sg9	or	sg7	or
7	or	obs	or	sg7
8	obs	sg9	obs	sg9

[3]Test loss from 1/16 to 1/22

[4]Number 3 stands for the number of features.

3 Result 2 (Exp-2)

Table 8: Evaluation of LSTM trained with positional encoding.

LSTM	Test loss	LSTM-3[5]	Test loss	Improvement
ew3	0.0158±0.0004	ew3	0.0149±0.0001	5.70%
ew2	0.0161±0.0000	ew2	0.0150±0.0003	6.83%
ew4	0.0163±0.0003	ew4	0.0152±0.0002	6.75%
sg5	0.0166±0.0001	sg7	0.0155±0.0003	13.89%
obs	0.0175±0.0001	sg5	0.0156±0.0001	11.86%
or	0.0177±0.0002	or	0.0156±0.0002	6.02%
sg7	0.0180±0.0002	sg9	0.0160±0.0005	14.89%
sg9	0.0188±0.0002	obs	0.0164±0.0003	6.29%

[5]Number 3 stands for the number of features.

4 Result 3 (Exp-3)

4.1 Colour in baseline performance

Table 9: Evaluation of each baseline approach for colour forecasting.

Rank	Model-Dataset	Test loss*	valid loss
1	LSTM-ew4	0.0136±0.0003	**0.7515±0.0310**(3)
2	LSTM-ew3	0.0138±0.0001	0.7547±0.0057
3	LSTM-ew2	0.0138±0.0001	0.8011±0.0131
4	GRU-ew3	0.0140±0.0003	0.8068±0.0070
5	GRU-ew2	0.0142±0.0001	0.8330±0.0104
6	LSTM-sg9	0.0143±0.0005	**0.7137±0.0216**(1)
7	GRU-ew4	0.0143±0.0001	0.7694±0.0071
8	RNN-ew3	0.0144±0.0002	0.8492±0.0371
9	RNN-sg9	0.0147±0.0003	0.8363±0.0125
10	RNN-ew4	0.0147±0.0001	0.8476±0.0238
11	LSTM-obs	0.0148±0.0003	0.9744±0.0124
12	GRU-obs	0.0149±0.0003	0.9927±0.0076
13	RNN-ew2	0.0150±0.0002	0.9083±0.0202
14	GRU-sg9	0.0151±0.0001	0.7575±0.0253
15	RNN-sg7	0.0158±0.0001	0.8755±0.0249
16	RNN-sg5	0.0158±0.0001	0.8846±0.0180
17	GRU-sg7	0.0159±0.0005	0.7791±0.0152
18	GRU-sg5	0.0160±0.0004	0.8080±0.0210
19	RNN-obs	0.0160±0.0001	1.0623±0.0394
20	LSTM-sg7	0.0161±0.0003	**0.7439±0.0364**(2)

Table 10: Evaluation of LSTM trained with positional encoding.

LSTM	Test loss	LSTM-3[6]	Test loss	Improvement
ew4	0.0136±0.0003	sg9	0.0120±0.0007	16.08%
ew3	0.0138±0.0001	ew2	0.0132±0.0004	4.35%
ew2	0.0138±0.0001	ew3	0.0134±0.0004	2.90%
sg9	0.0143±0.0005	ew4	0.0135±0.0003	0.74%
obs	0.0148±0.0003	obs	0.0135±0.0001	8.78%
sg7	0.0161±0.0003	sg7	0.0143±0.0003	11.18%
sg5	0.0168±0.0005	sg5	0.0144±0.0002	14.29%

[6]Number 3 stands for the number of features.

5 Conclusion

Table 11: Influence of pre-processing method on NH_3N and colour forecasting models trained with LSTM.

Rank	NH_3N	NH_3N + pos[7]	Colour	Colour + pos
1	**ew3**	**ew3**	**ew4**	sg9
2	**ew2**	**ew2**	**ew3**	**ew2**
3	**ew4**	**ew4**	**ew2**	**ew3**
4	sg5	sg7	sg9	**ew4**
5	obs	sg5	obs	obs
6	or	or	sg7	sg7
7	sg7	sg9	sg5	sg5
8	sg9	obs		

[7] Postional encoding.