DEEP LEARNING FINAL PROJECT

Arlette Michelle Perez Espinoza
Miray Yüce
Mostafa Hussein

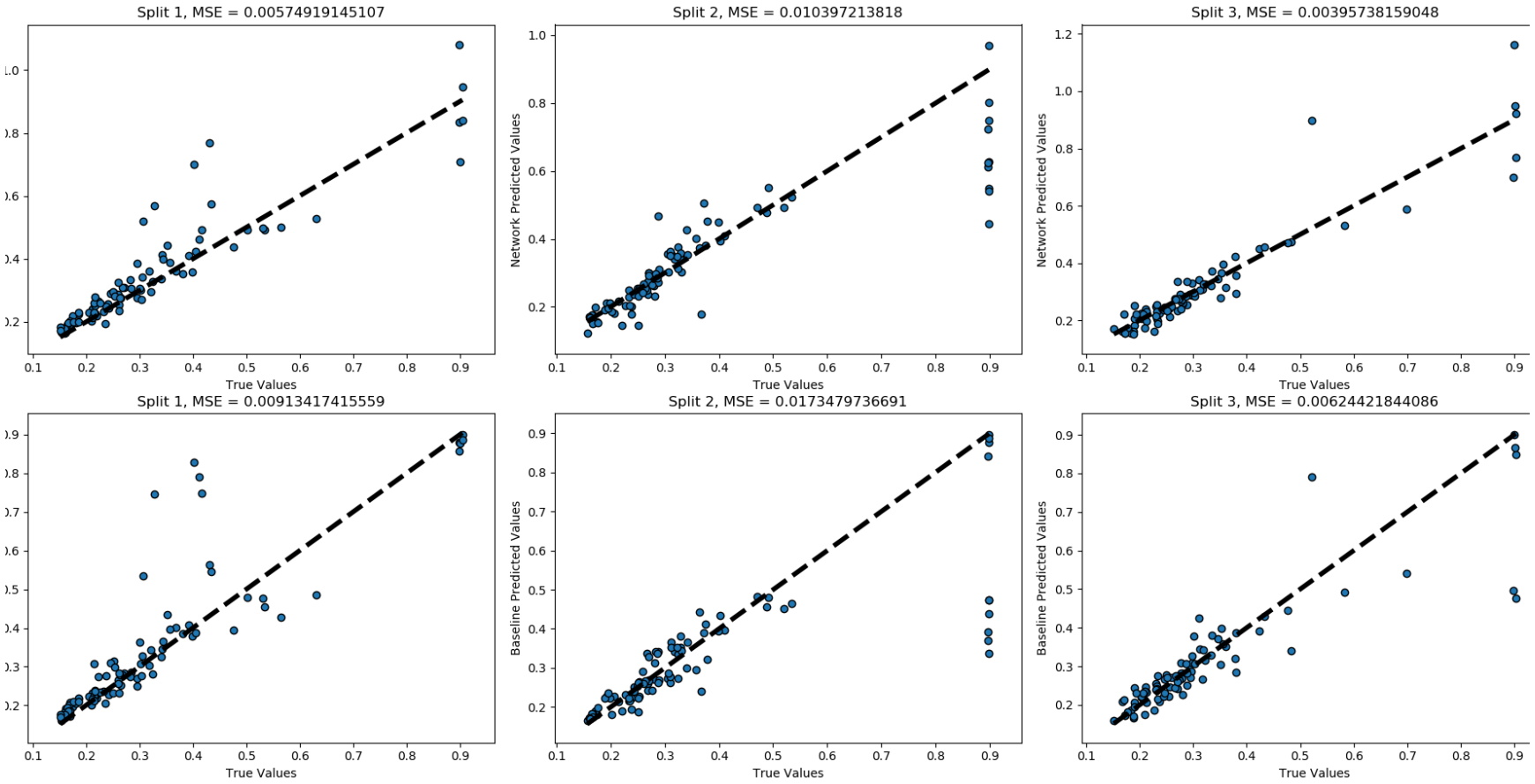
Task 1: Prediction of the final error using MLP

- Two hidden layers with 64 units each, and ReLU activation functions.
- Kernel initialization: random uniform (-0.05, 0.05)
- Bias initialization: constant 0.1
- Randomized hyperparameter search is done for 10 models for each regularization approach.

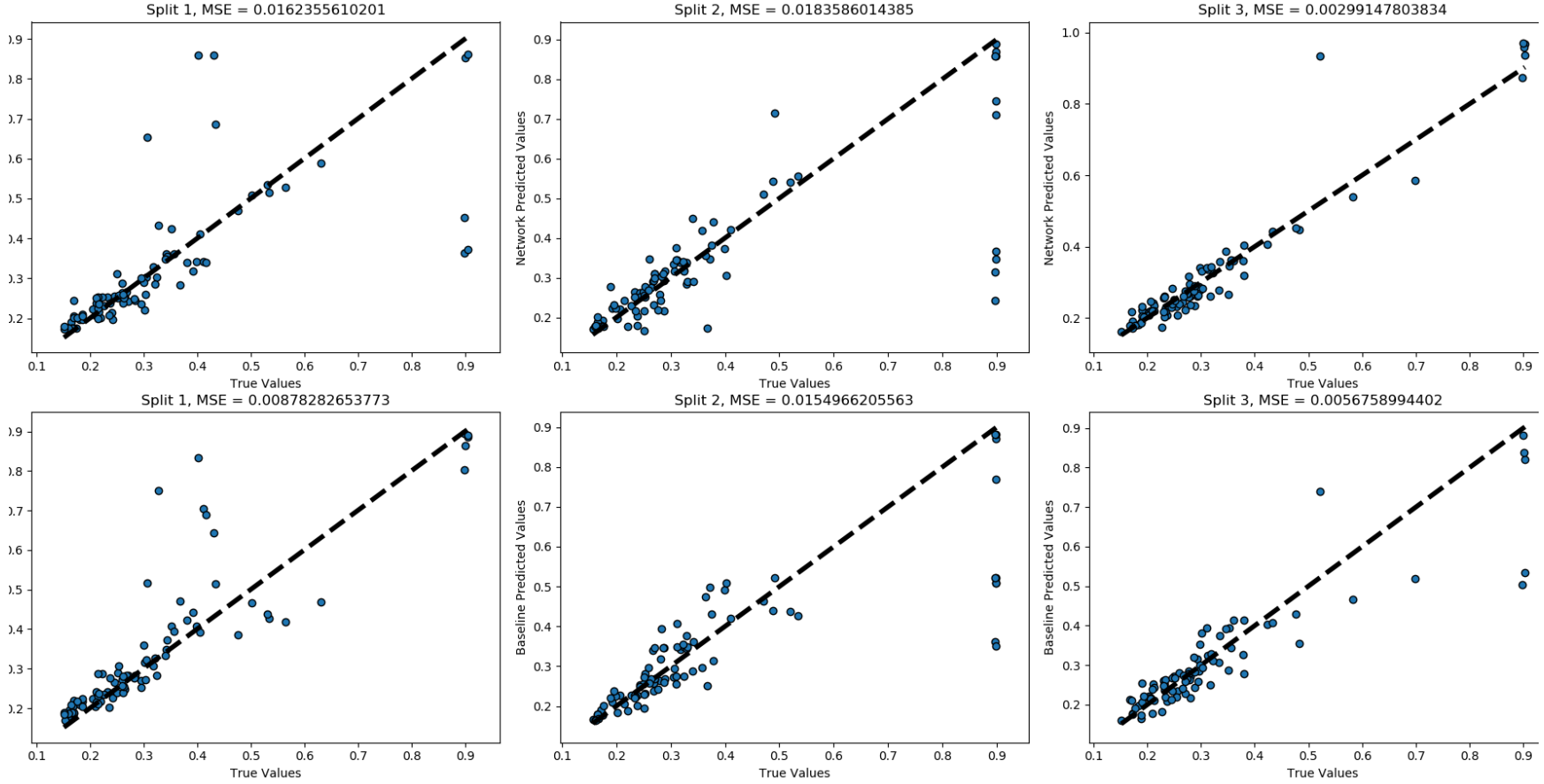
Parameter	Min Value	Max Value	Log
Batch size	32	512	No
Learning rate	10^{-6}	10^{0}	Yes
# units in layer 1	16	1024	Yes
# units in layer 2	16	1024	Yes
# units in layer 3	16	1024	Yes

MLP and Random Forest results for raw data

	Network					Baseline			
Experiment	LR	Alpha	Batch size	Noise	Mean MSE of 3 splits	Max depth	# Estimators	Min leaf	Mean MSE of 3 splits
w/out reg.	0.001	0	265	0	**0.0067**	32	32	2	0.0109
L1	1e-5	0.0001	64	0	0.007	32	16	4	0.0101
L2	1e-5	0.0001	64	0	0.009	32	16	4	0.0101
Noise	1e-4	1e-5	8	0.02	**0.012**	16	32	4	0.0099
Decay	1e-4 – 1e-6	0	128	0	0.0083	16	8	4	0.0101
L1-Noise	1e-4	0.001	128	0.02	0.0076	8	16	4	0.0101
L2-Noise	1e-5	1e-5	128	0.02	0.0084	8	16	4	0.0101
L1-Decay	0.01 – 1e-6	1e-4	265	0	0.0087	8	16	4	0.0101
L2-Decay	0.001 – 1e-6	1e-5	265	0	0.0090	8	16	4	0.0101
Noise-Decay	0.001 – 1e-6	0	128	0.02	0.0082	16	8	4	0.0101
L1-Noise-Decay	0.001 – 1e-6	1e-5	265	0.02	0.0071	8	16	4	0.0101
L2-Noise-Decay	0.01 – 1e-6	1e-4	265	0.02	0.0079	8	16	4	0.0101

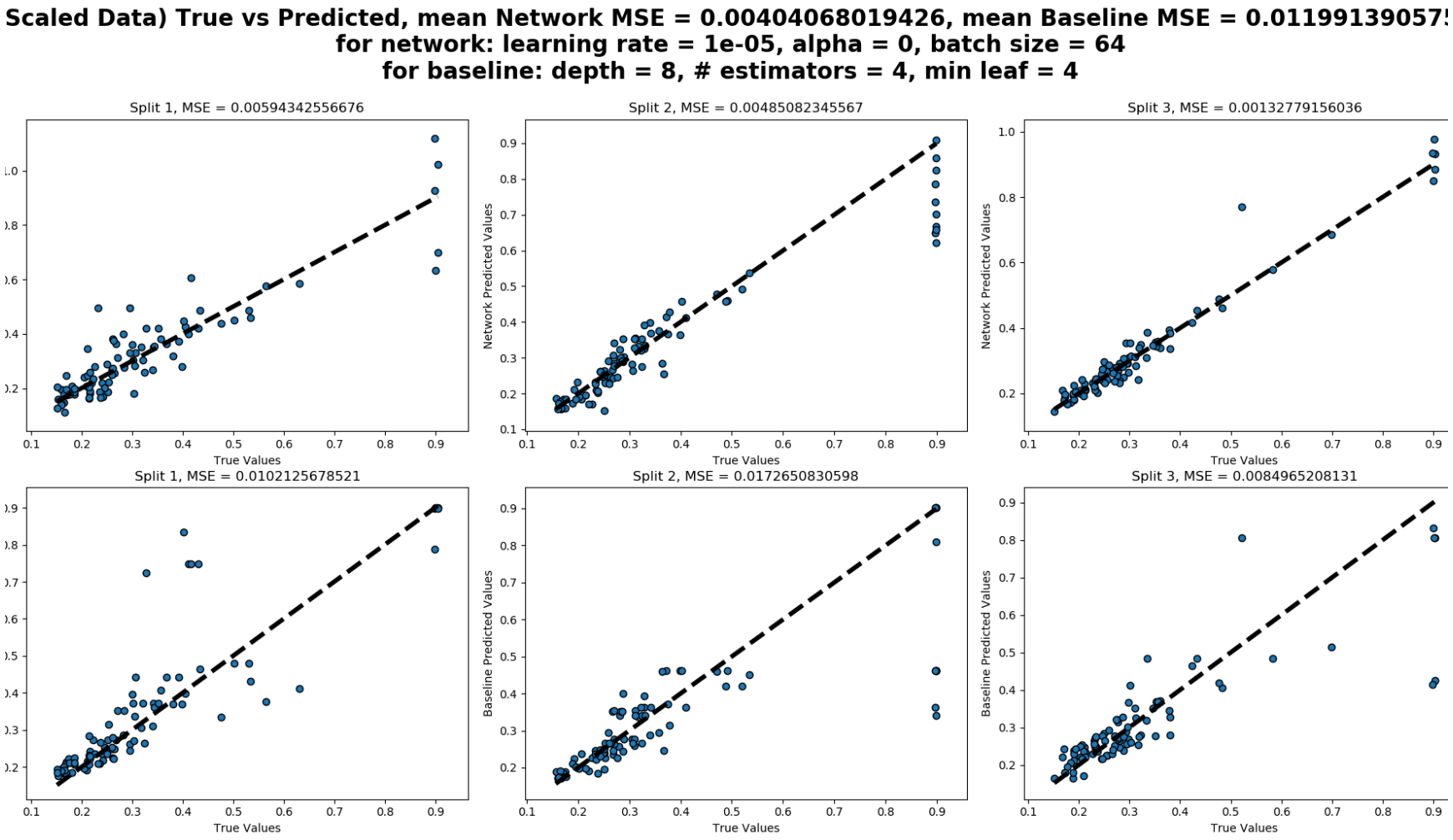

(Raw Data) True vs Predicted, mean Network MSE = 0.00670126228651, mean Baseline MSE = 0.0109087887552
for network: learning rate = 0.001, alpha = 0, batch size = 265
for baseline: depth = 32, # estimators = 32, min leaf = 2
Split 1, MSE = 0.00574919145107
Split 2, MSE = 0.010397213818
Split 3, MSE = 0.00395738159048
Network Predicted Values
True Values
Split 1, MSE = 0.00913417415559
Split 2, MSE = 0.0173479736691
Split 3, MSE = 0.00624421844086
Baseline Predicted Values
True Values

(Raw Data) True vs Predicted, mean Network MSE = 0.0125285468323, mean Baseline MSE = 0.00998511551141
for network: learning rate = 0.0001, alpha = 1e-05, batch size = 8
for baseline: depth = 16, # estimators = 32, min leaf = 4

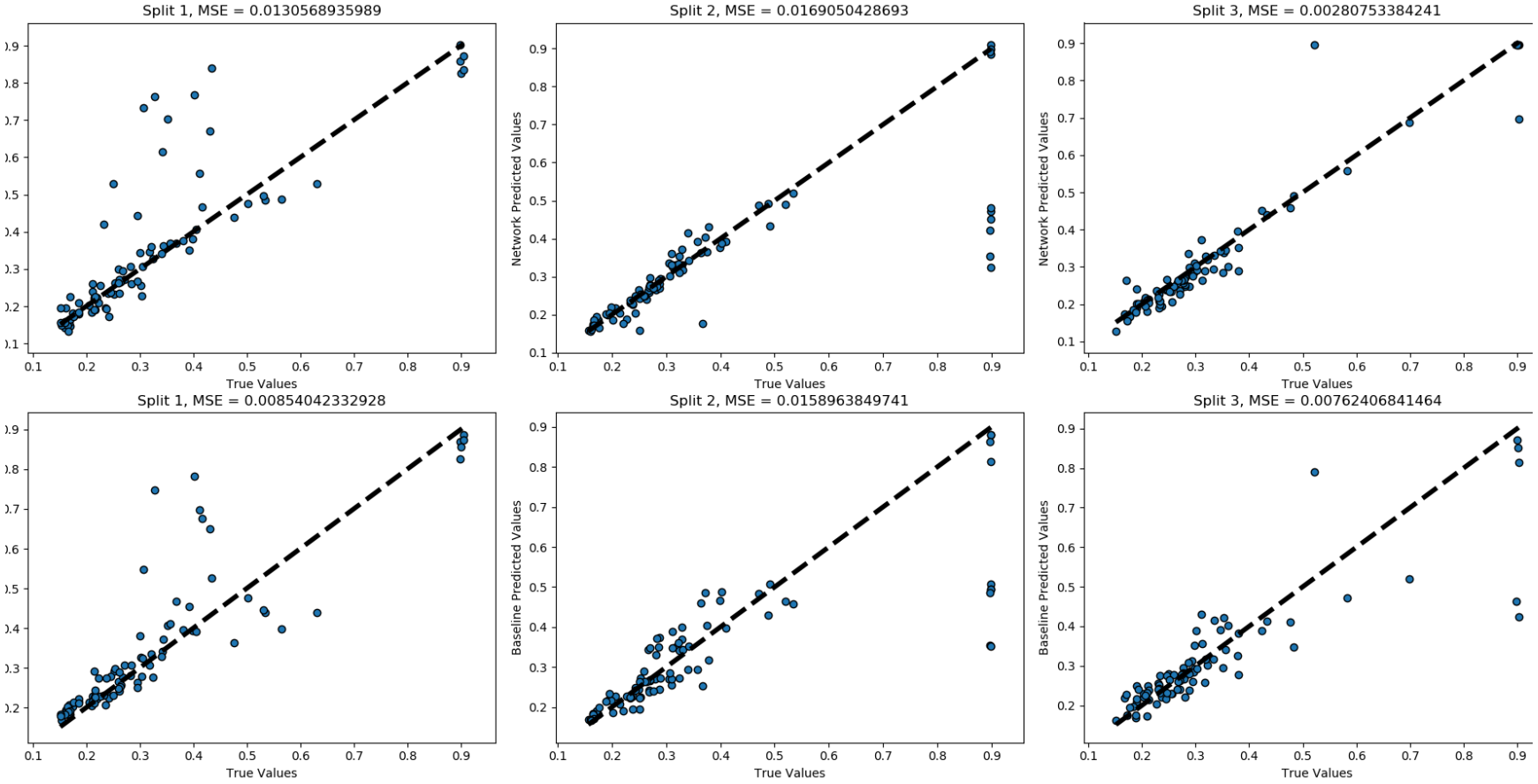


MLP and Random Forest results for scaled data

	Network					Baseline			
	LR	Alpha	Batch size	Noise	Mean MSE of 3 splits	Max depth	# Estimators	Min leaf	Mean MSE of 3 splits
w/out reg.	1e-5	0	64	0	**0.0040**	8	4	4	0.0119
L1	1e-5	0.0001	64	0	**0.0109**	16	32	4	0.0106
L2	1e-5	0.0001	0.0080	0	0.009	16	32	4	0.0106
Noise	1e-4	1e-5	8	0.02	0.0084	16	32	4	0.0099
Decay	1e-4 – 1e-6	0	128	0	0.0051	16	4	4	0.0123
L1-Noise	1e-5	1e-5	32	0.02	0.0080	32	32	4	0.0106
L2-Noise	1e-6	1e-4	64	0.02	0.0078	32	32	4	0.0106
L1-Decay	0.01 – 1e-6	1e-5	8	0	0.0102	8	16	4	0.0105
L2-Decay	0.001 – 1e-6	1e-5	265	0	0.0087	8	16	4	0.0105
Noise-Decay	1e-4 – 1e-6	0	128	0.02	0.0072	16	4	4	0.0123
L1-Noise-Decay	0.001 – 1e-6	1e-5	265	0.02	0.0059	8	16	4	0.0105
L2-Noise-Decay	0.001 – 1e-6	1e-5	265	0.02	0.0055	8	16	4	0.0105


Scaled Data) True vs Predicted, mean Network MSE = 0.00404068019426, mean Baseline MSE = 0.011991390575
for network: learning rate = 1e-05, alpha = 0, batch size = 64
for baseline: depth = 8, # estimators = 4, min leaf = 4
Split 1, MSE = 0.00594342556676
Split 2, MSE = 0.00485082345567
Split 3, MSE = 0.00132779156036
Split 1, MSE = 0.0102125678521
Split 2, MSE = 0.0172650830598
Split 3, MSE = 0.0084965208131
True Values
Network Predicted Values
Baseline Predicted Values

(Scaled Data) True vs Predicted, mean Network MSE = 0.0109231567702, mean Baseline MSE = 0.010686958906
for network: learning rate = 1e-05, alpha = 0.0001, batch size = 64
for baseline: depth = 16, # estimators = 32, min leaf = 4



Task 2: Extrapolation of learning curves

Architecture

- 2 LSTM layers, 64 units each
- 2 Dense layers, 64 units each, ReLu activation

Regularizations:

- Gaussian Noise of 0.02 between layers, except before the output layer
- L2 regularization of kernels for all layers

Optimizations:

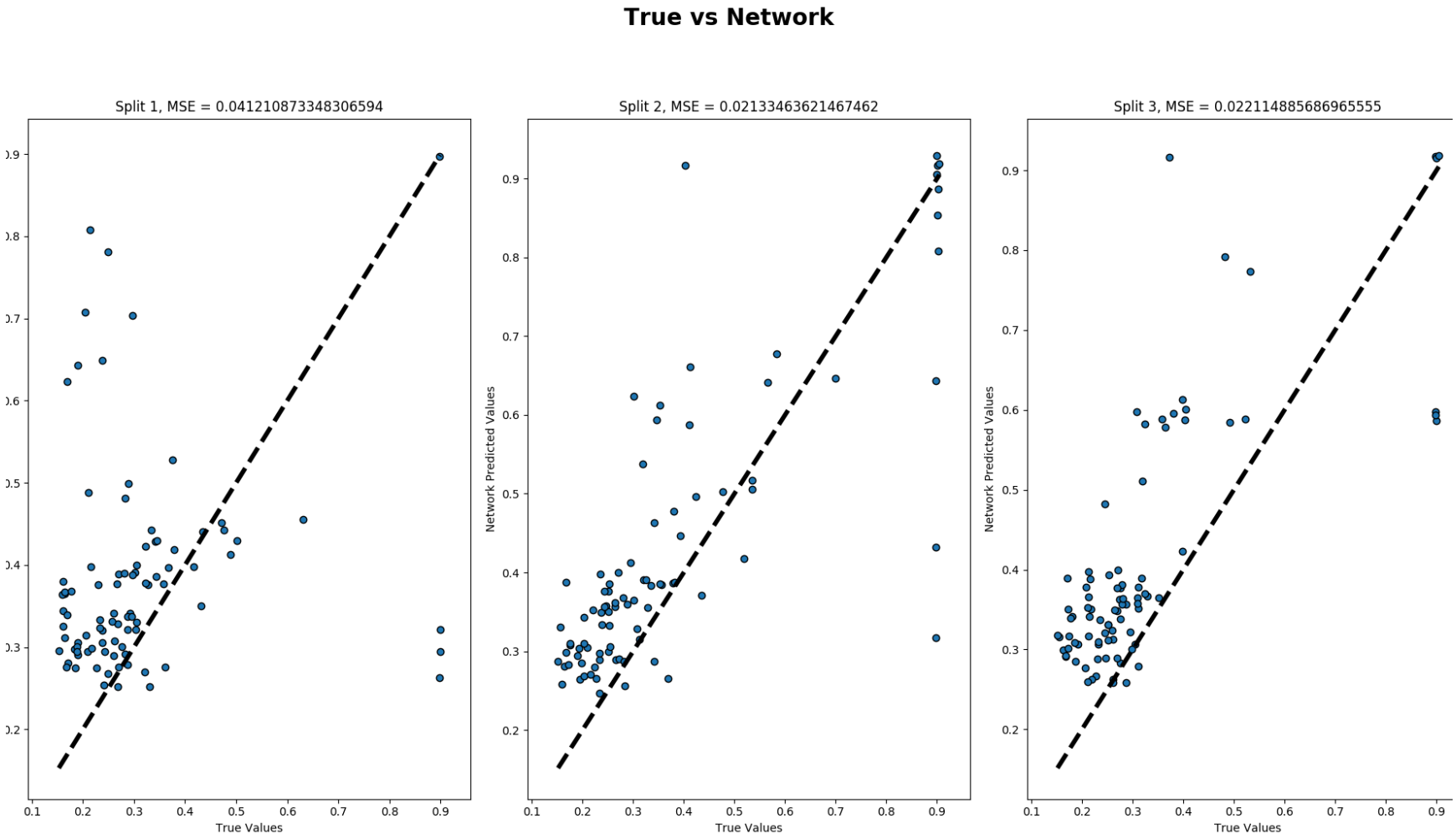
- Adam
- Kernel initialization for all layers: random uniform [0.01, 0.05]
- Bias initialization for all layers: constant 0.1

Results of task 2: RNN's and baselines

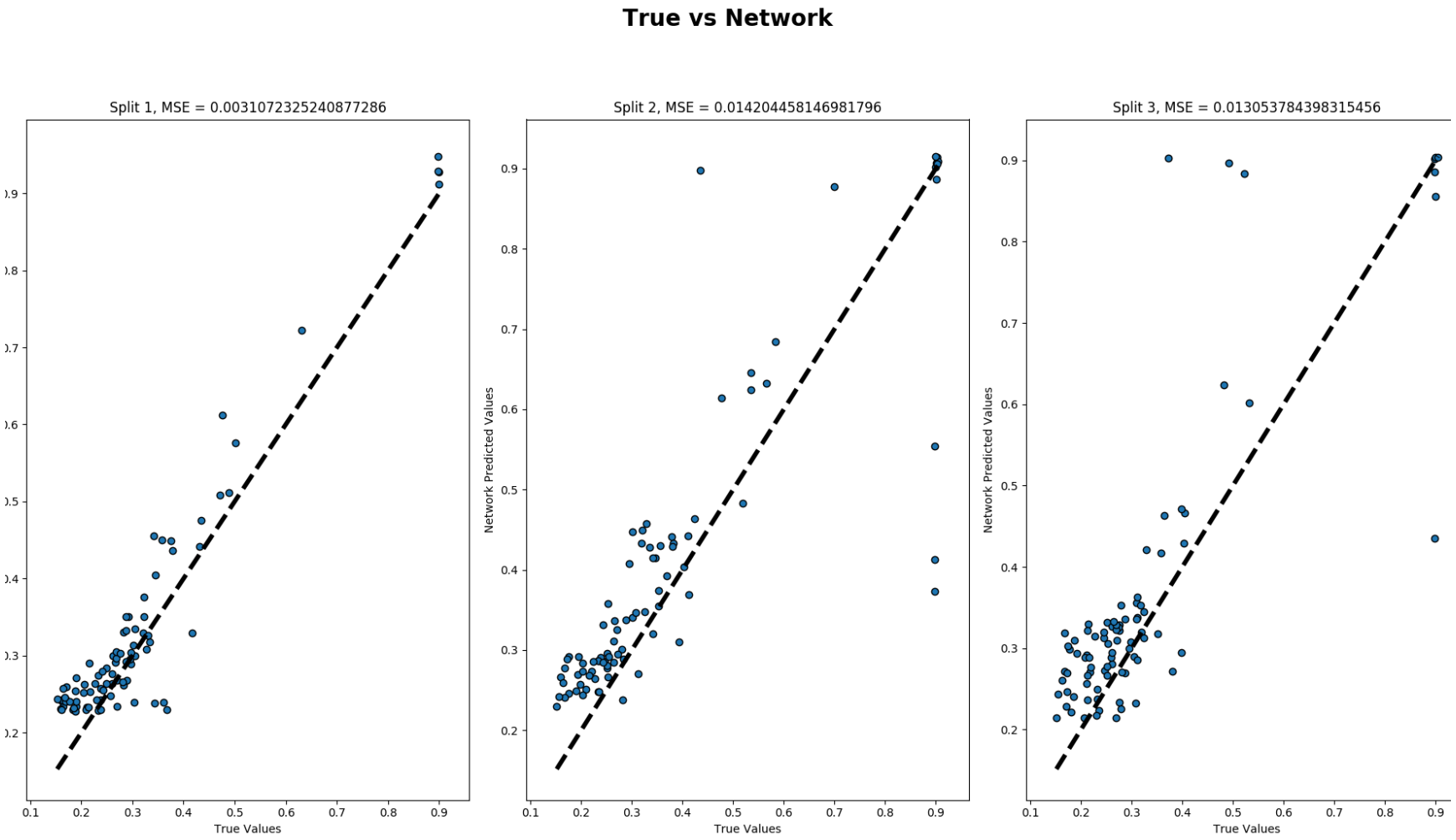
	LR	Alpha	Pred 5 MSE	Pred 10 MSE	Pred 20 MSE	Pred 30 MSE
Train 5	0.001 – 1e-6	1e-7	0.0282	0.0267	0.0257	0.0272
Train 10	1e-4 – 1e-6	1e-5	0.0228	0.01011	0.0272	0.0264
Train 20	0.001 – 1e-6	1e-6	0.0279	0.0260	0.0068	0.0283
Random	1e-4 – 1e-7	1e-5	0.0089	0.0066	0.0056	0.0046

	Max depth	# Estimators	Min leaf	MSE
Train 5	32	16	1	0.0015
Train 10	32	32	4	0.0010
Train 20	32	16	1	0.0004
Last 4 point prediction	32	16	8	0.0003

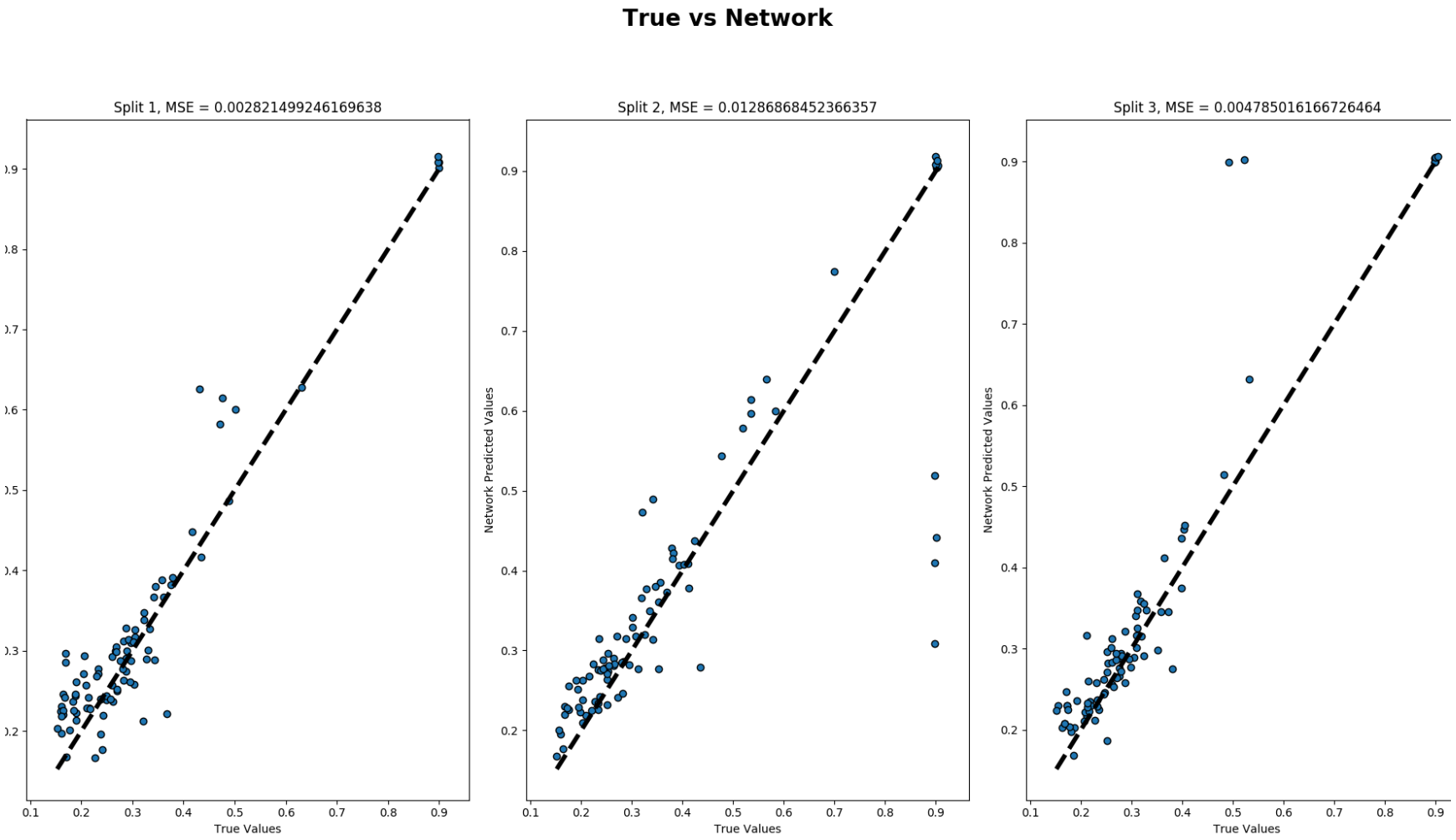
RNN trained with fixed input length: 5

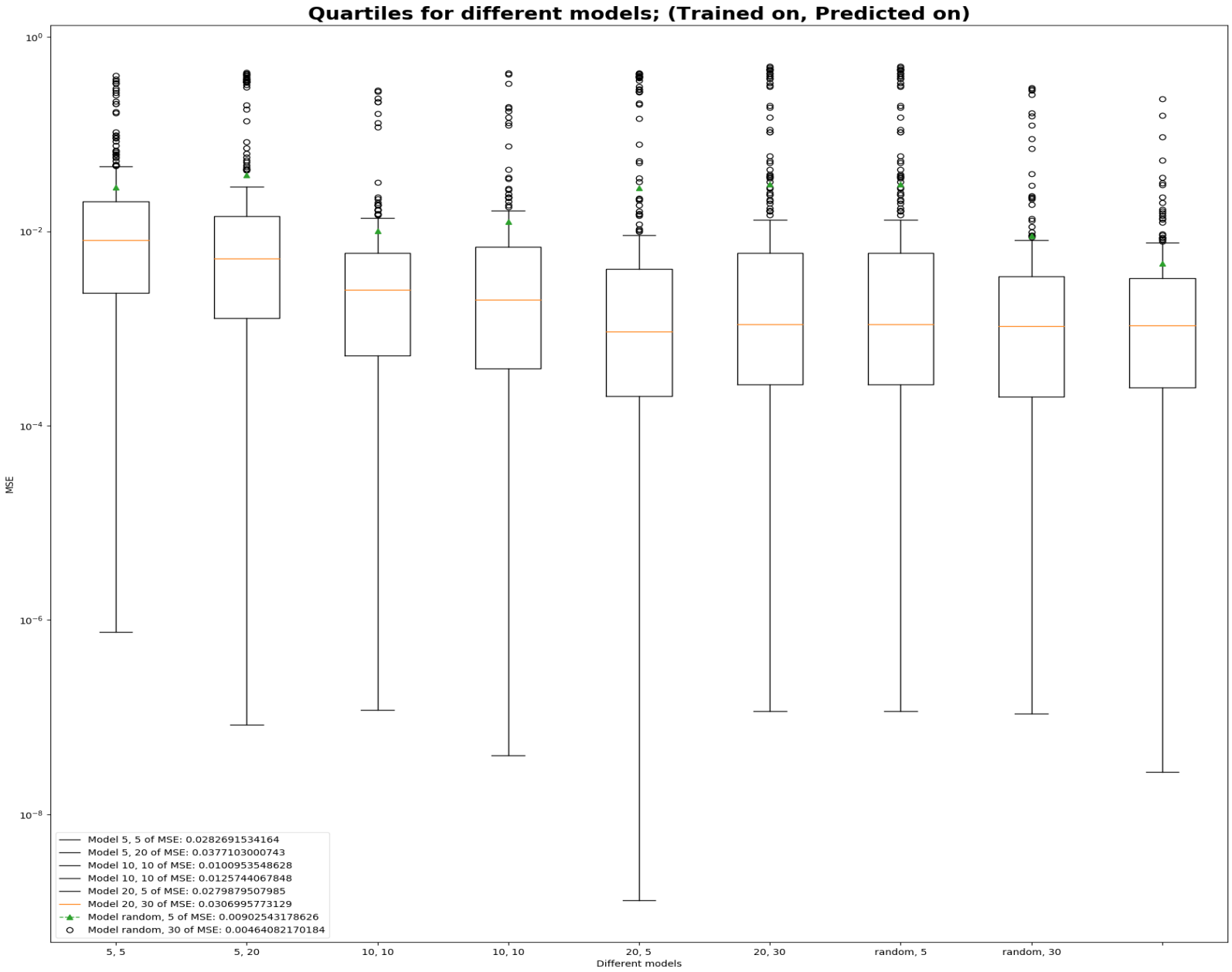


RNN trained with fixed input length: 10

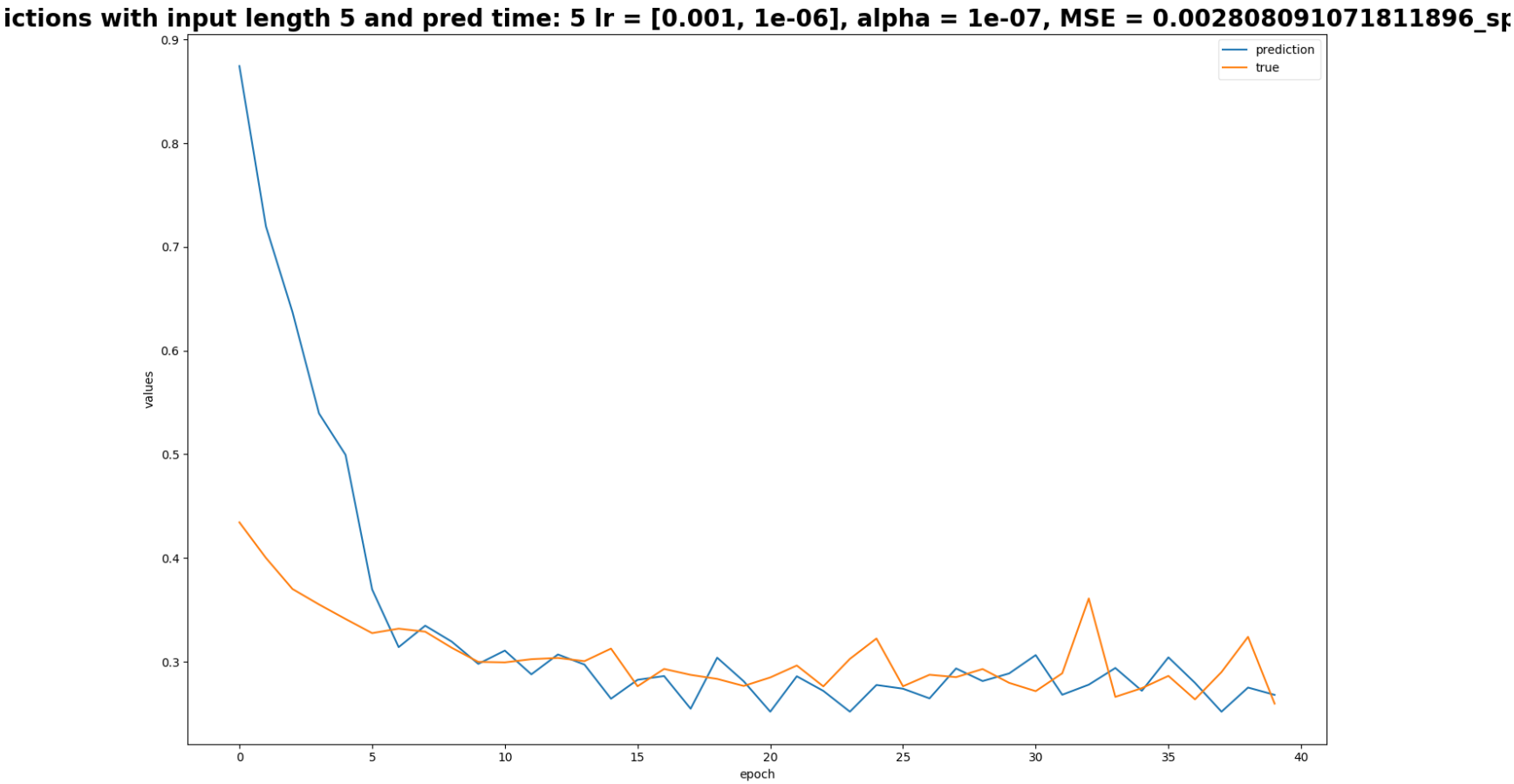


RNN trained with fixed input length: 20

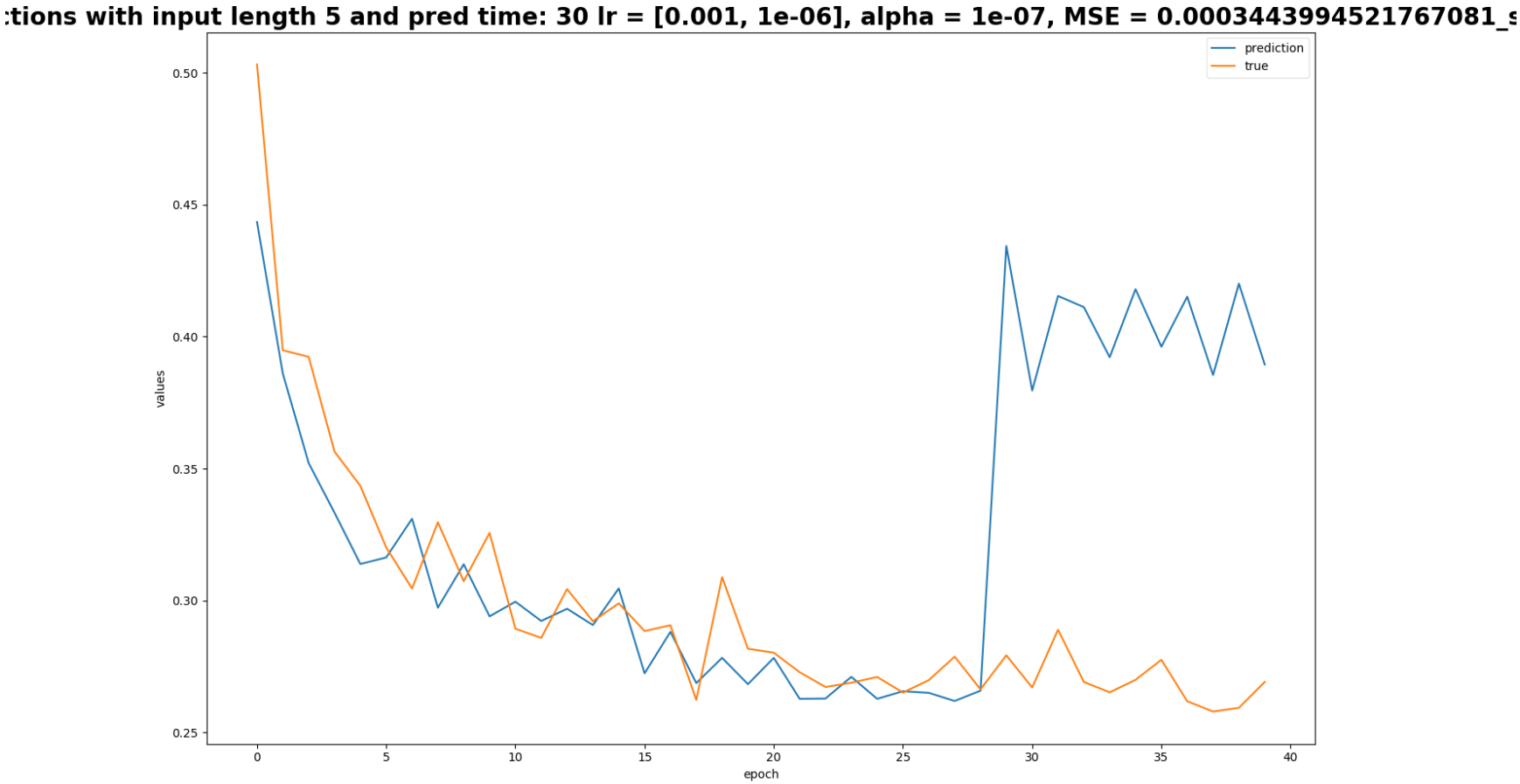



Quartiles for different models; (Trained on, Predicted on)
MSE
Different models
5, 5
5, 20
10, 10
10, 10
20, 5
20, 30
random, 5
random, 30
Model 5, 5 of MSE: 0.0282691534164
Model 5, 20 of MSE: 0.0377103000743
Model 10, 10 of MSE: 0.0100953548628
Model 10, 10 of MSE: 0.0125744067848
Model 20, 5 of MSE: 0.0279879507985
Model 20, 30 of MSE: 0.0306995773129
Model random, 5 of MSE: 0.00902543178626
Model random, 30 of MSE: 0.00464082170184

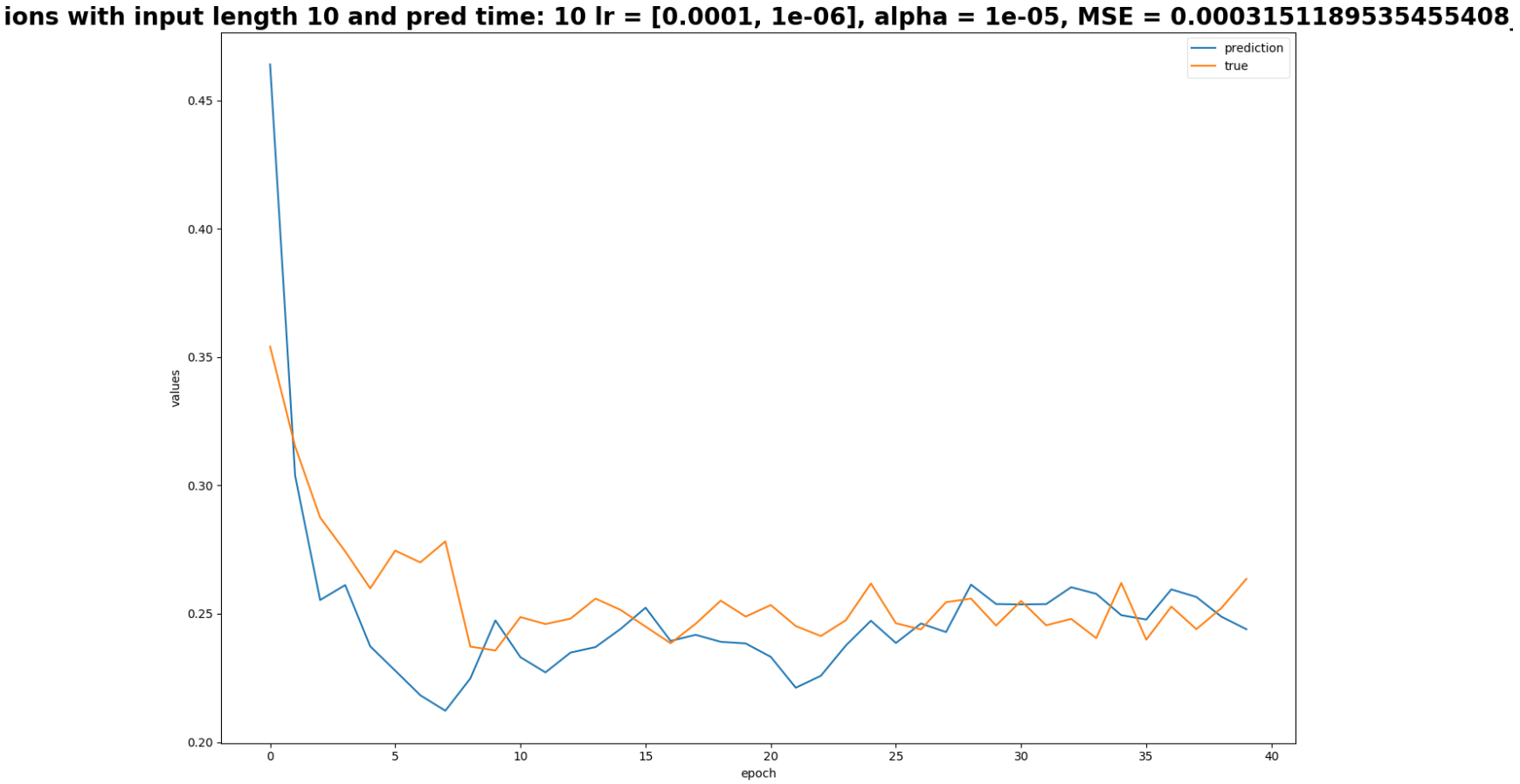
Train on 5, predictions of split 1



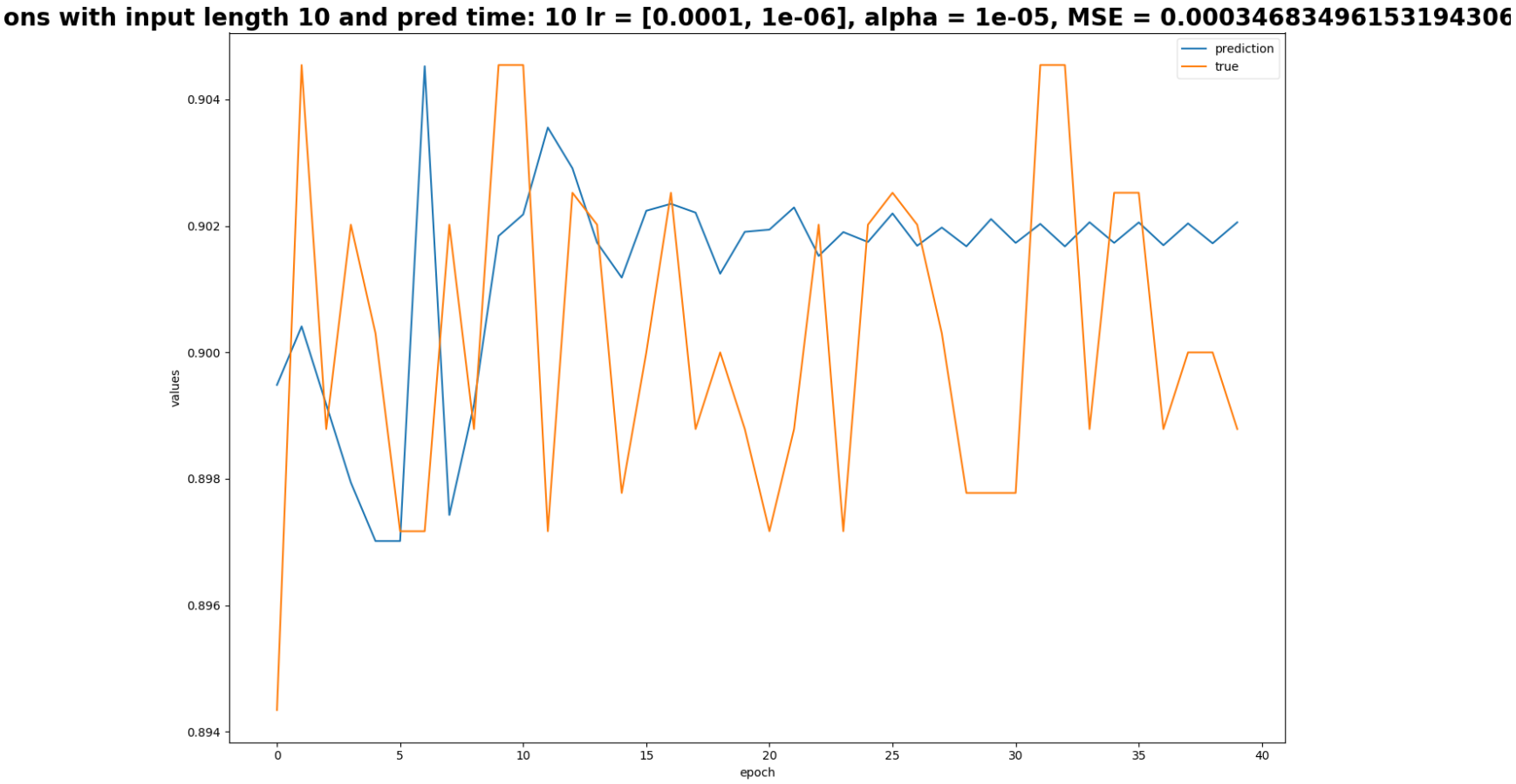
Train on 5, predictions of split 3



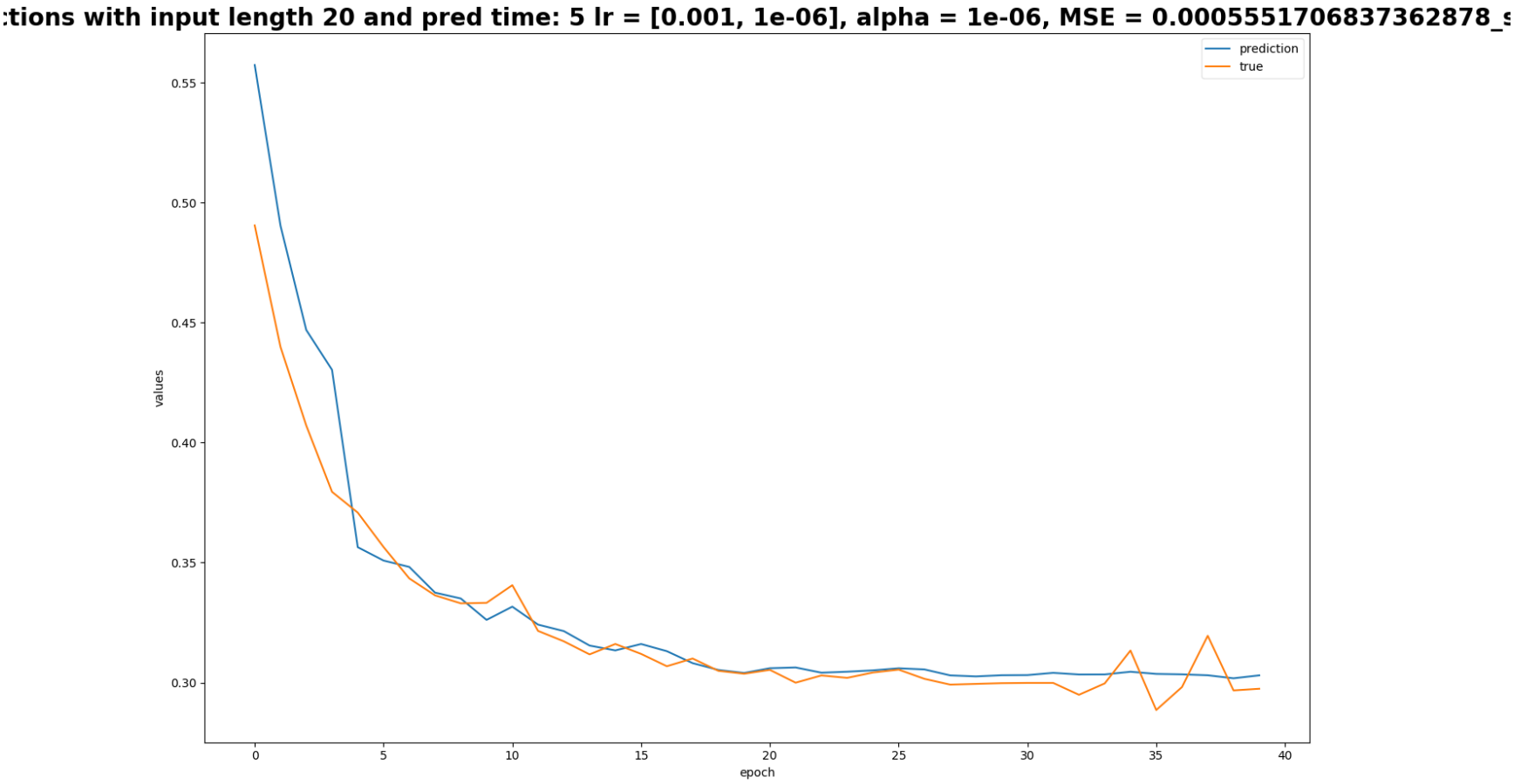
Train on 10, predictions of split 2



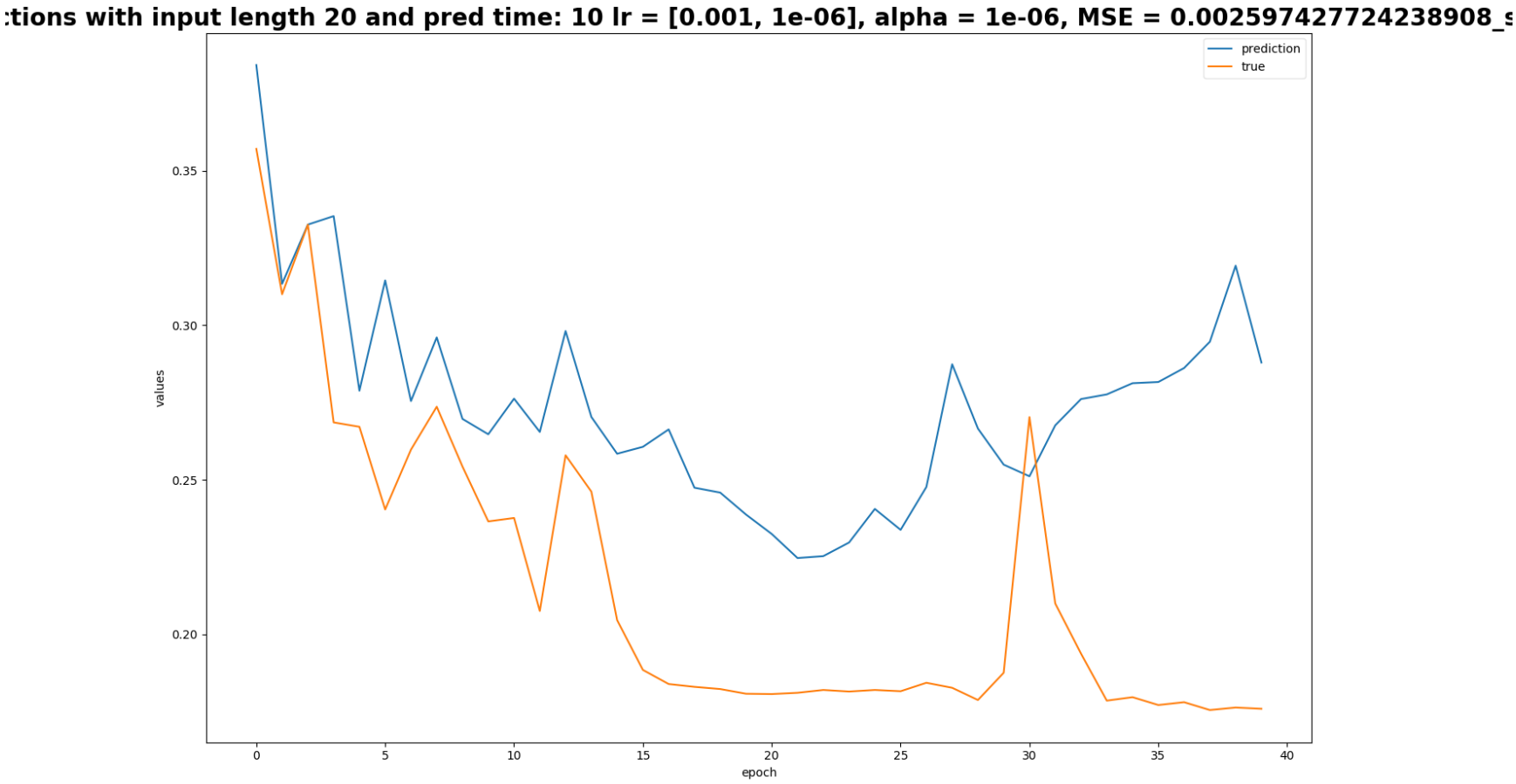
Train on 10, predictions of split 3



Train on 20, predictions of split 1

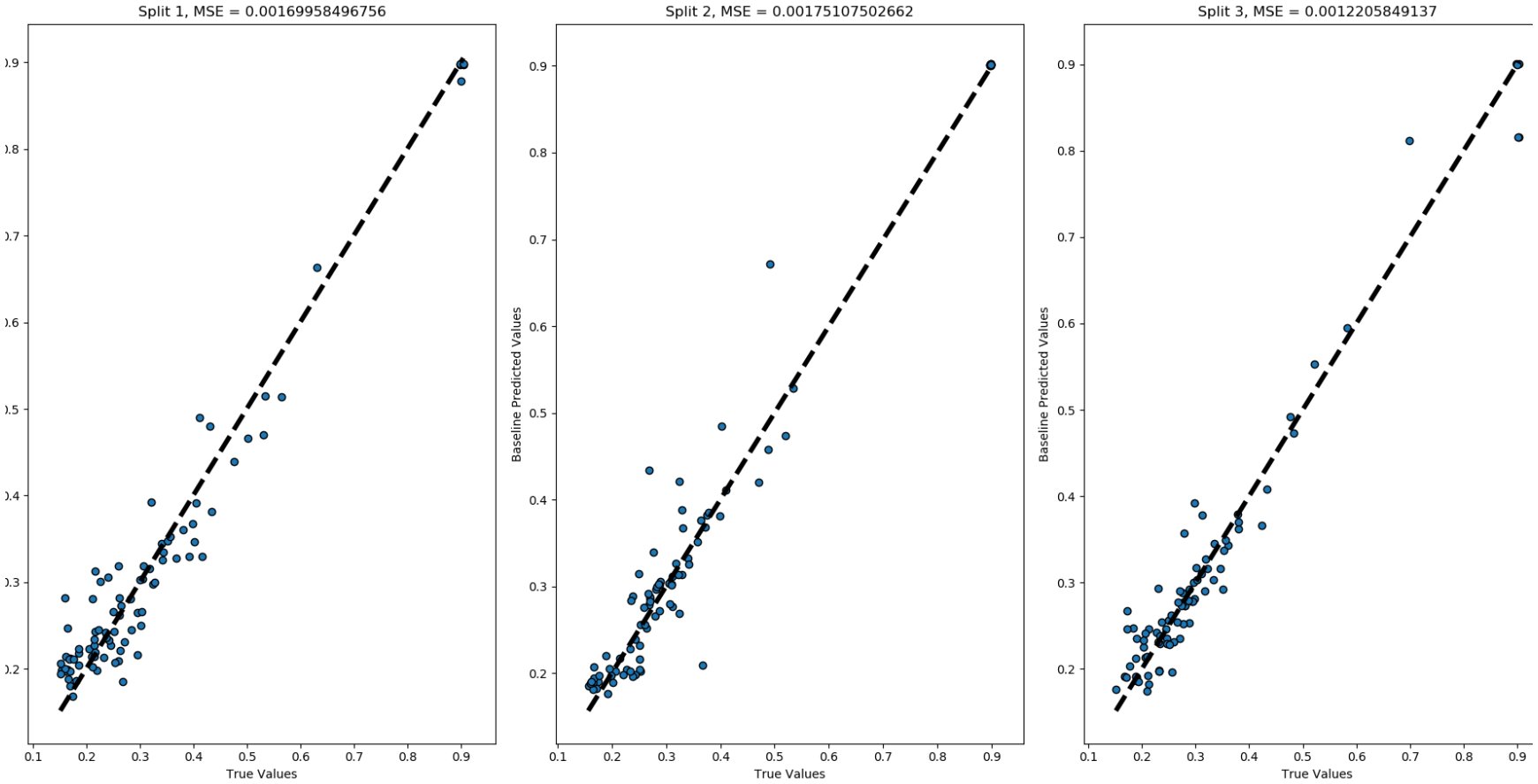


Train on 20, predictions of split 2



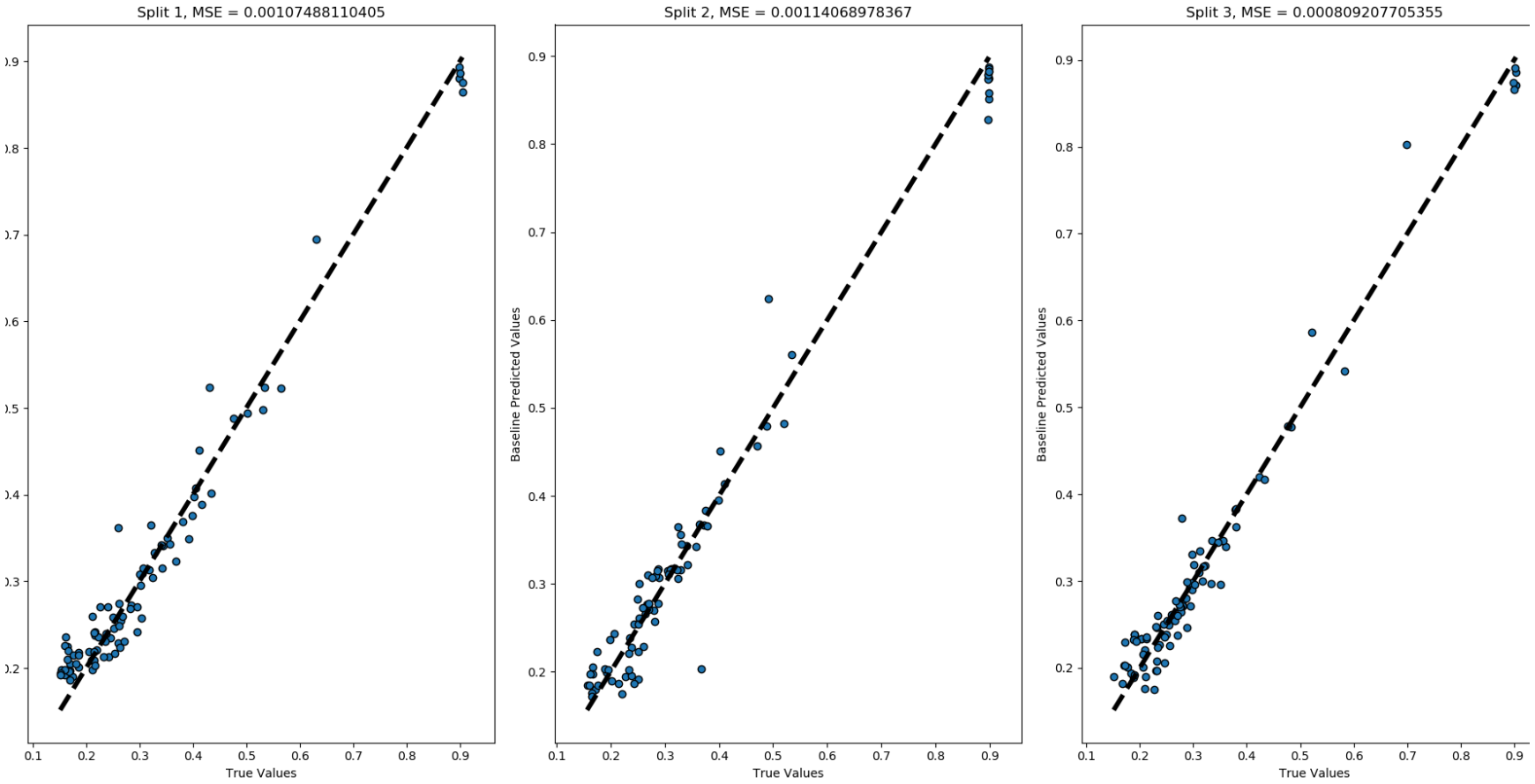
Baseline with fixed number of observations: 5

True vs Baseline with depth = 32, # estimators = 16, min leaf = 1, Mean MSE = 0.00155708163596



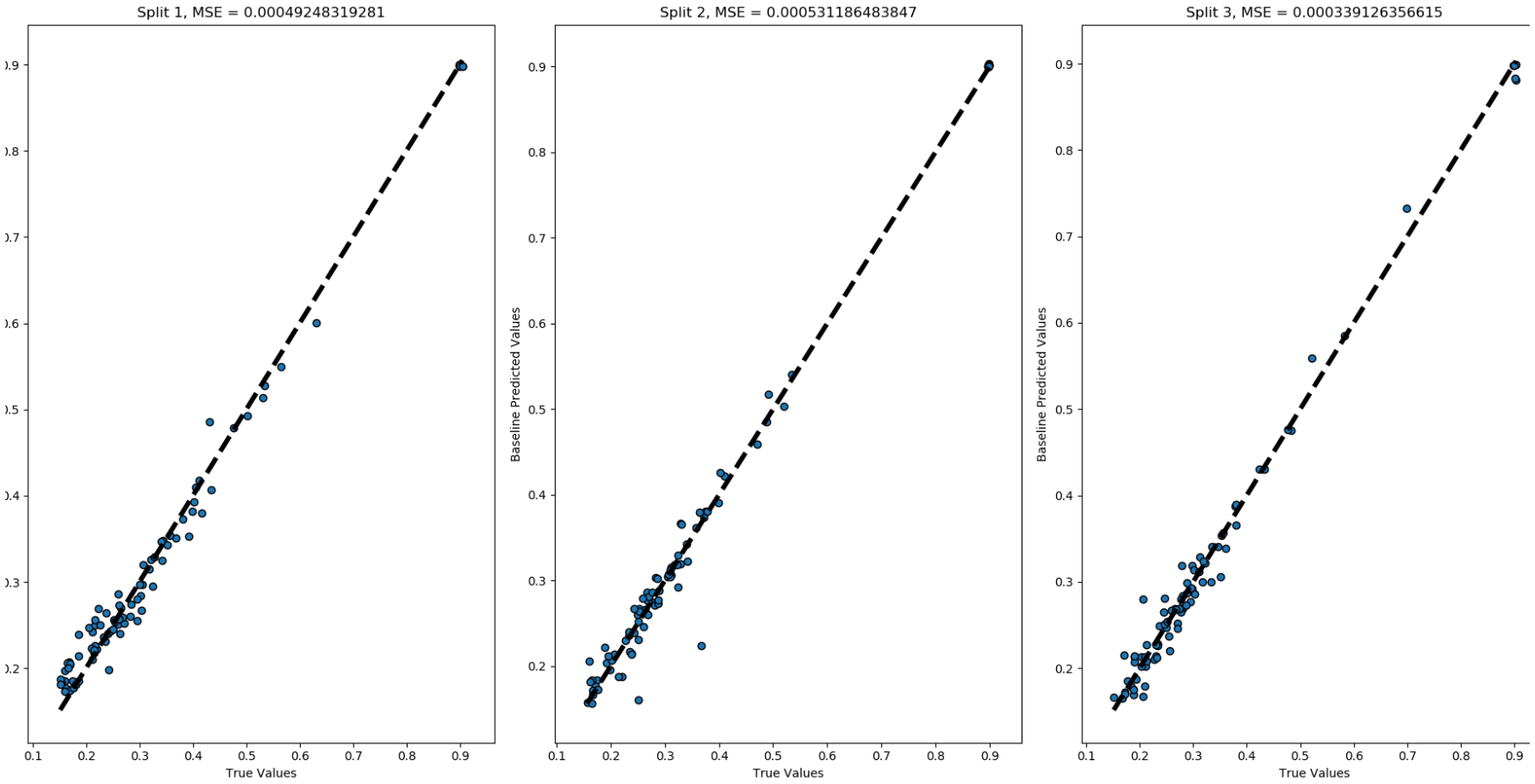
Baseline with fixed number of observations: 10

True vs Baseline with depth = 32, # estimators = 32, min leaf = 4, Mean MSE = 0.00100825953103

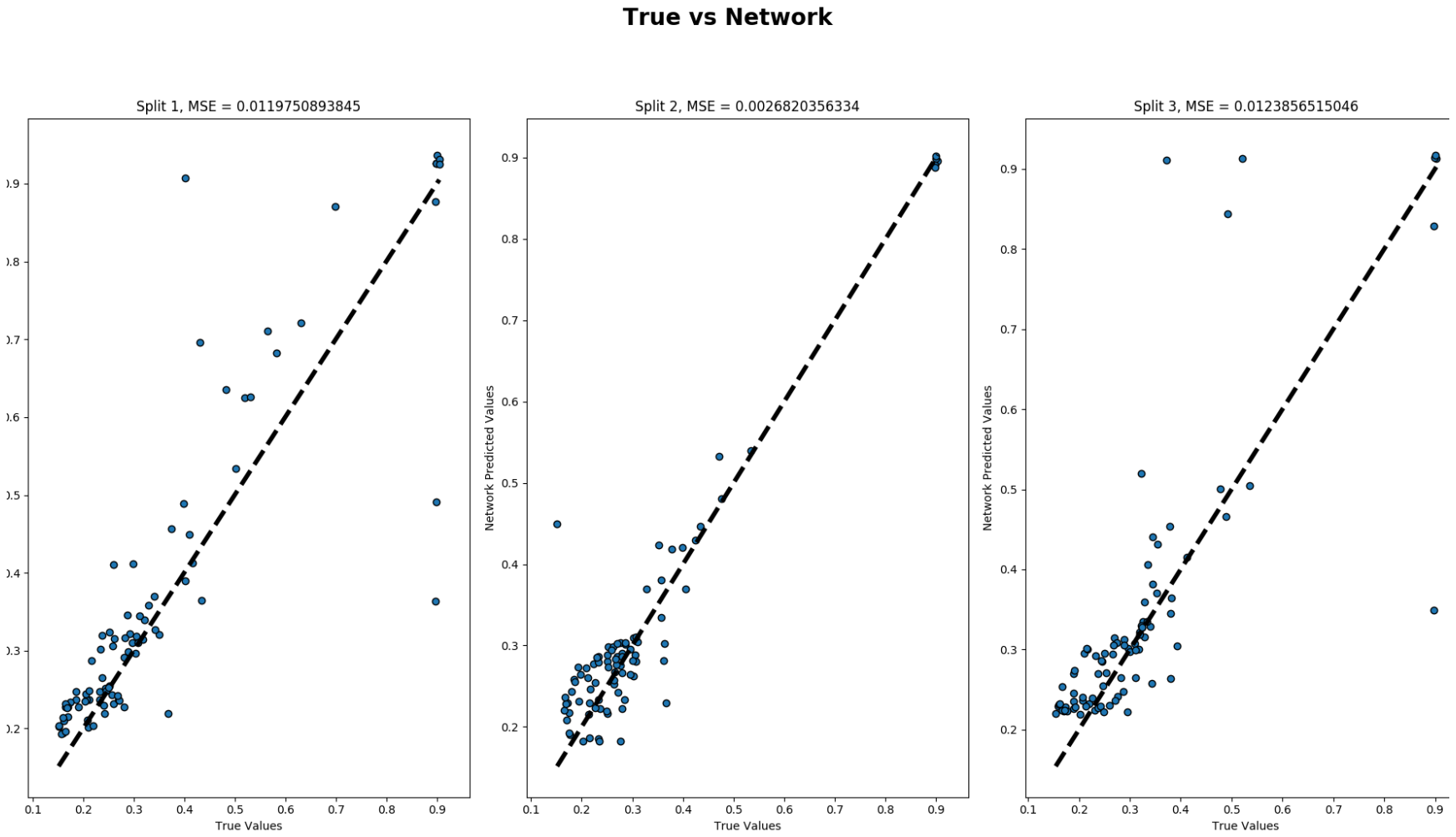


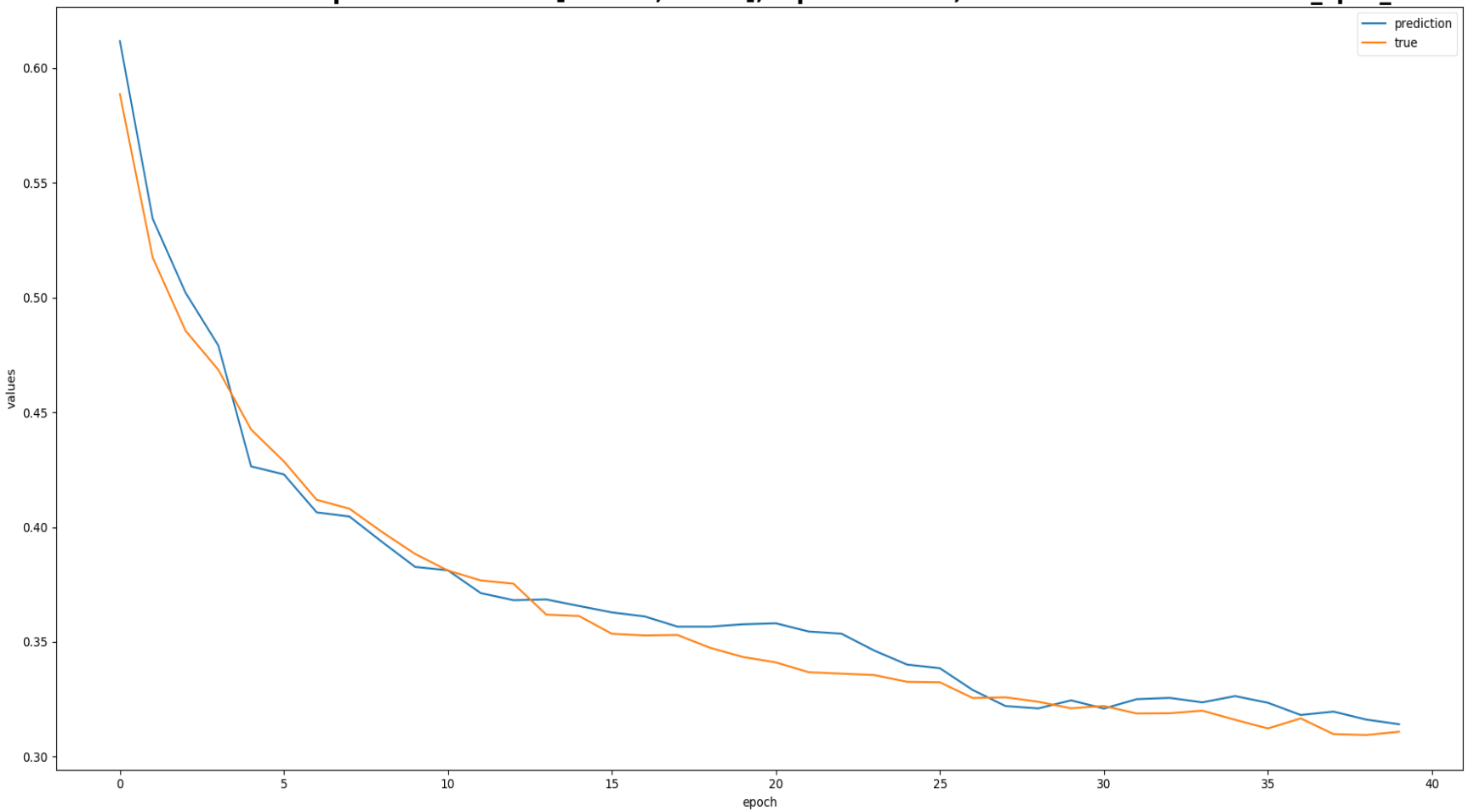
Baseline with fixed number of observations: 20

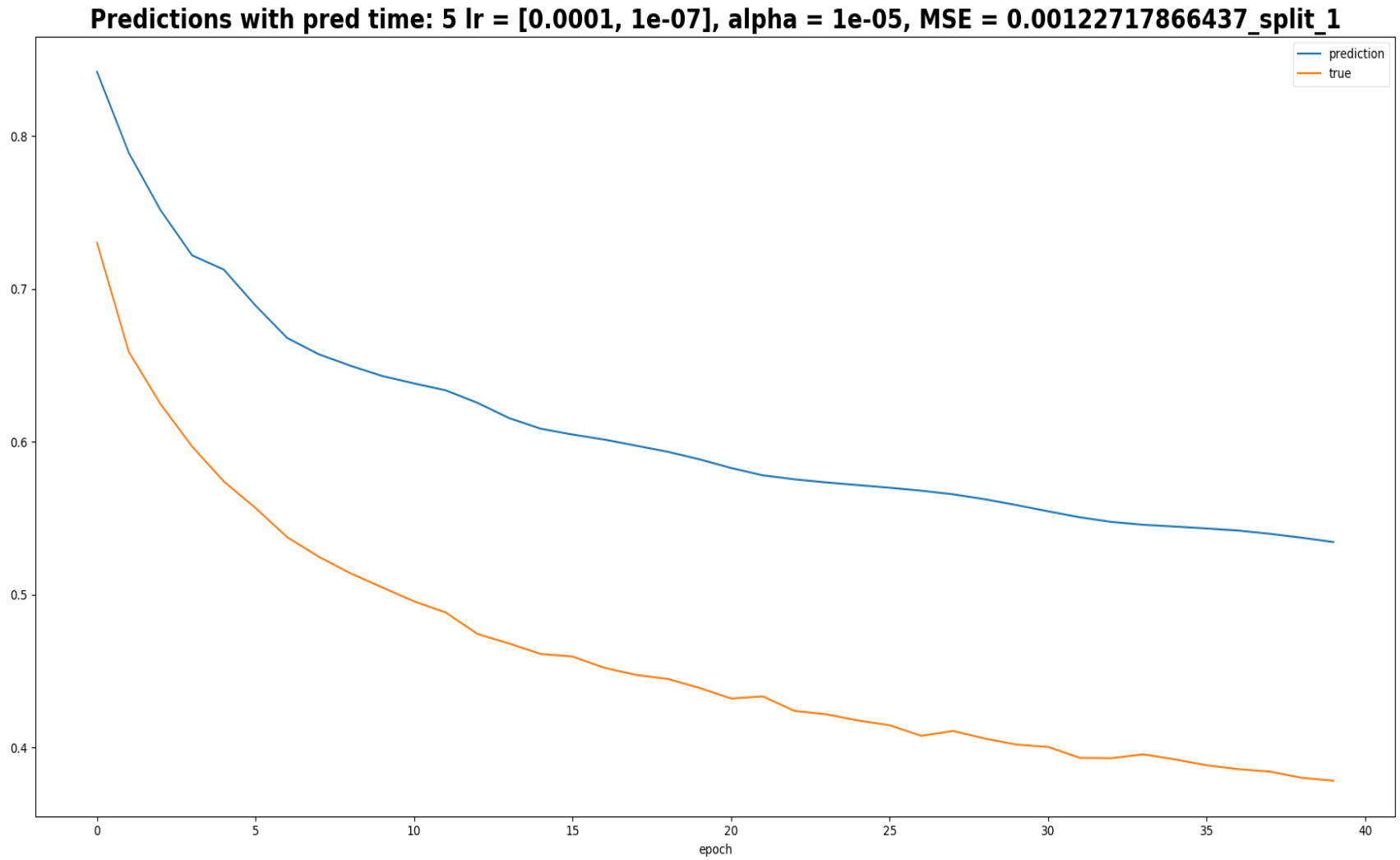
True vs Baseline with depth = 32, # estimators = 16, min leaf = 1, Mean MSE = 0.000454265344424



RNN trained with random input length, predictions after 5th epoch

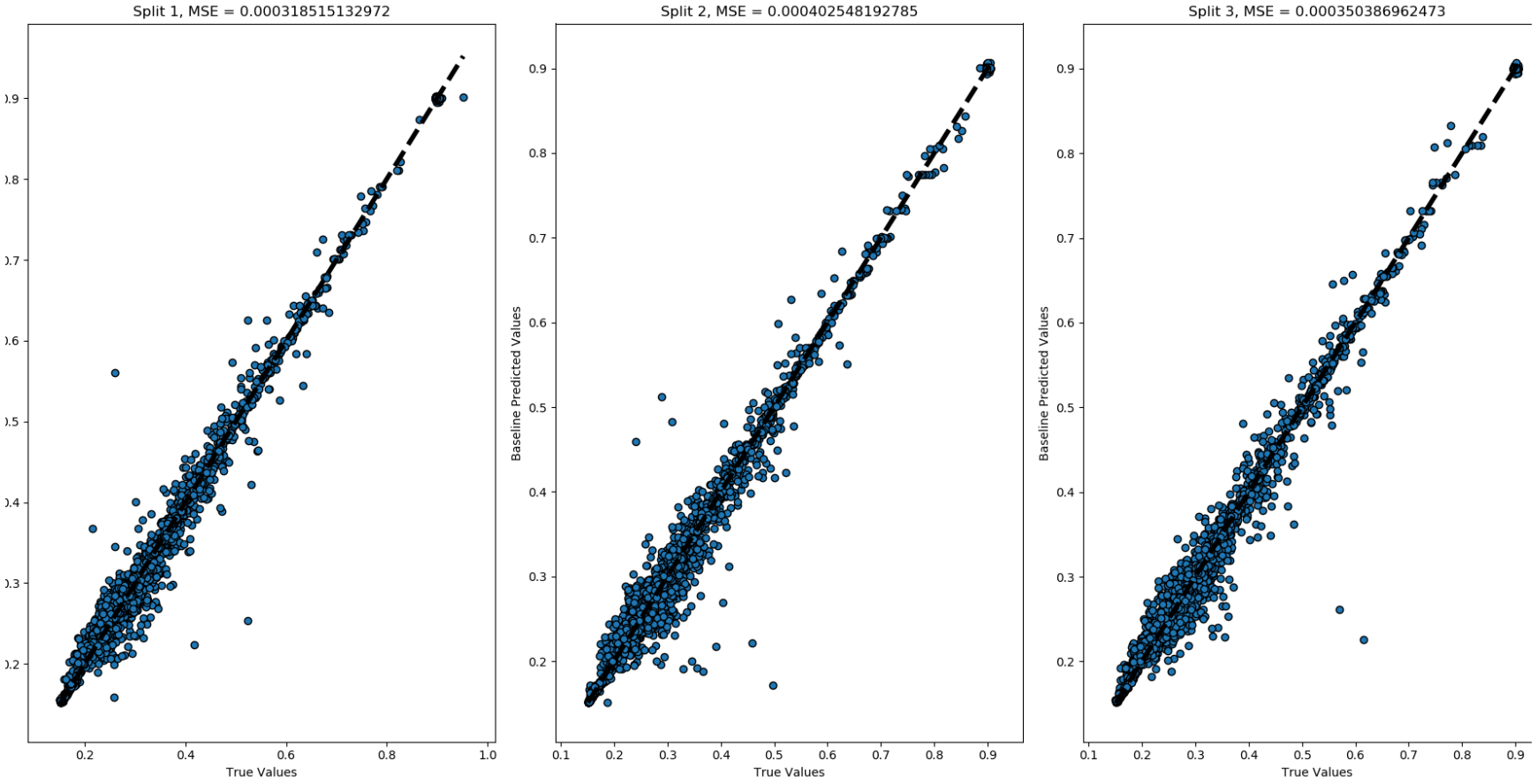



Predictions with pred time: 5 lr = [0.0001, 1e-07], alpha = 1e-05, MSE = 0.00122717866437_split_1
prediction
true
values
epoch
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0
5
10
15
20
25
30
35
40


Predictions with pred time: 5 lr = [0.0001, 1e-07], alpha = 1e-05, MSE = 0.00122717866437_split_1
prediction
true
0.8
0.7
0.6
0.5
0.4
0
5
10
15
20
25
30
35
40
epoch

Baseline to predict next point given previous four

True vs Baseline with depth = 32, # estimators = 16, min leaf = 8, Mean MSE = 0.000357150096076



Comparison between RNN's and last baseline

	LR	Alpha	Pred 5 MSE	Pred 10 MSE	Pred 20 MSE	Pred 30 MSE
Train 5	0.001 – 1e-6	1e-7	0.0282	0.0267	0.0257	0.0272
Train 10	1e-4 – 1e-6	1e-5	0.0228	0.01011	0.0272	0.0264
Train 20	0.001 – 1e-6	1e-6	0.0279	0.0260	0.0068	0.0283
Random	1e-4 – 1e-7	1e-5	0.0089	0.0066	0.0056	0.0046

	Max depth	# Estimators	Min leaf	MSE
Last 4 point prediction	32	16	8	0.0003