VRIJE UNIVERSITEIT BRUSSEL

Predicting Zero Point energies

3D CNN v0.1.2

Michiel Jacobs
04/02/2021

Run information

Model:

Author: Michiel Jacobs
Version: 0.1.2
Type: 3D CNN
Feature: Coulomb Matrix
Label: Zero point energy

Data:

Maximum heavy atoms: 20
Maximum molecule size: 62
Total molecules: 500

Tensorisation:

Positive dimensions: 0
Negative dimensions: 5

Test and train sets:

Split ratio: 0.8
Molecules for training: 400
Molecules for testing: 100

Neural Network

Network compile parameters:

Learningrate: 0.0001
Loss: mean_squared_error
Optimizer: Adam
Metrics: mean_absolute_error, mean_squared_error

Network fit parameters:

Batch size: 128
Epochs: 500
Validation split: 0.2
Shuffle data each epoch: True

Early stopping parameters:

Minimum change required: 0.0001
Epochs no change is allowed before stopping: 5
Restore best weights: True

Neural network Layer settings:

input shape: (6, 62, 62, 1)
kernel size: (1, 3, 3)
activation: relu
pool size: (2, 2, 2)
filters: 64
dropout: 0.2
dense units: 32
output shape: 1

NN summary

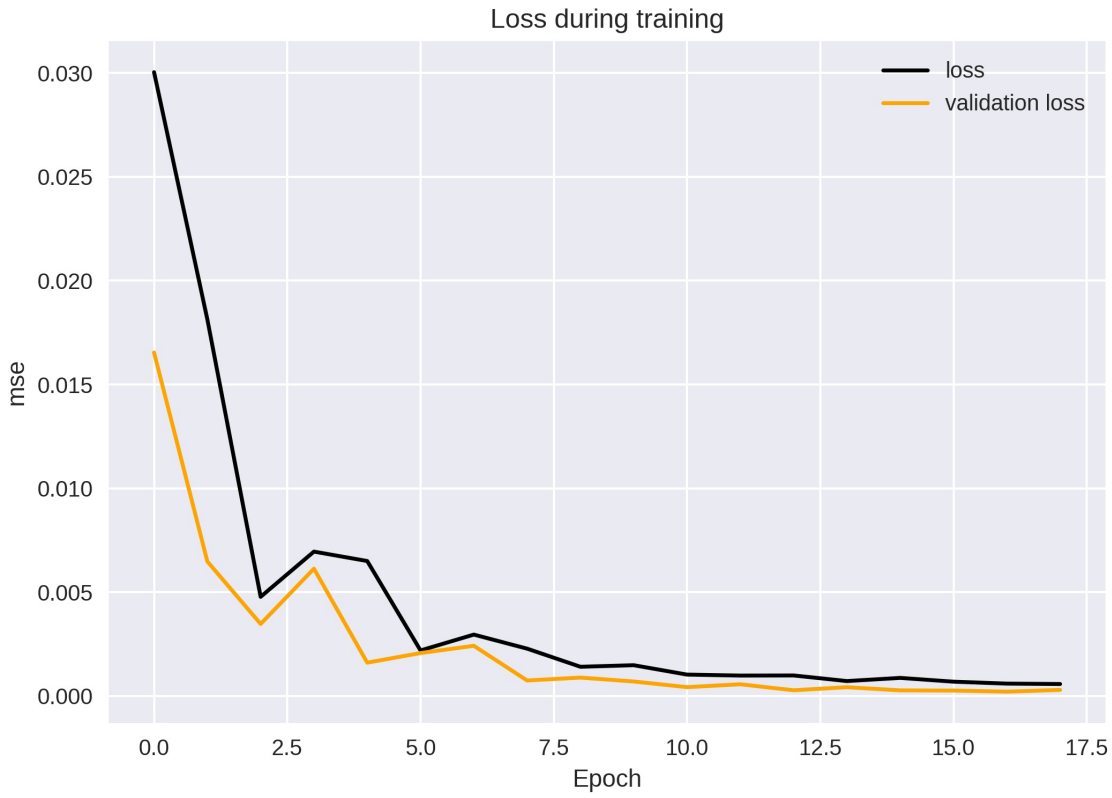
Model: "sequential"

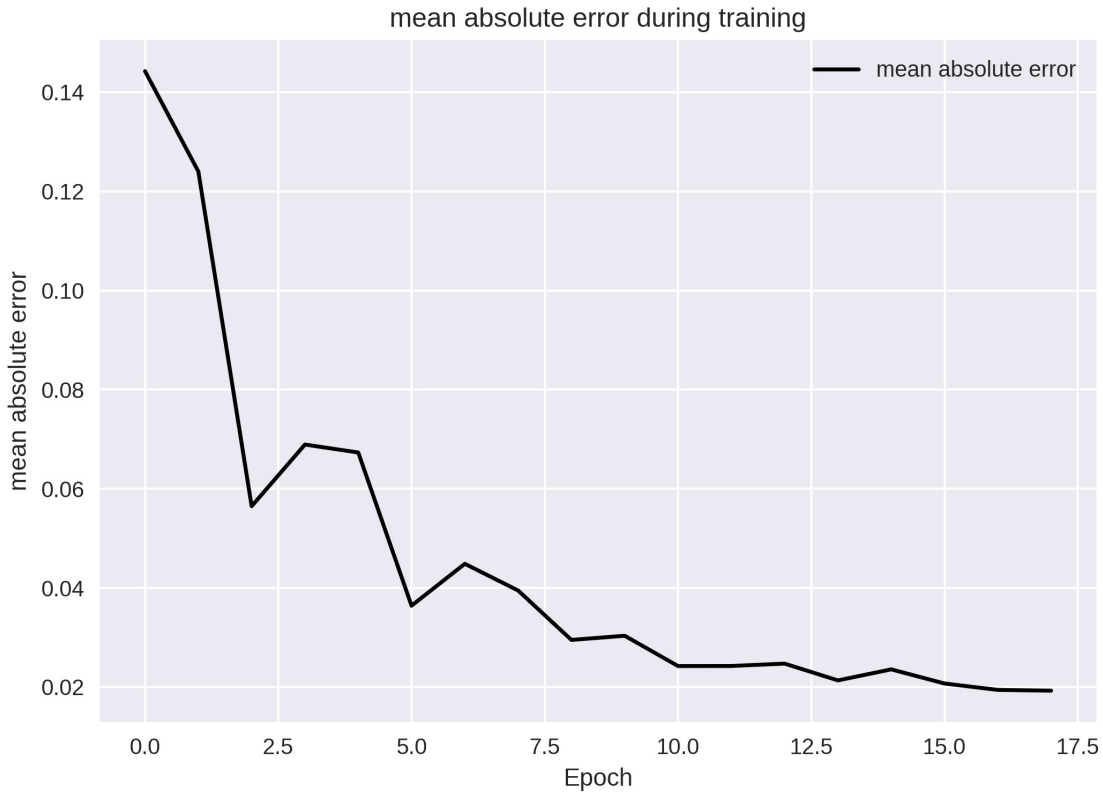
Layer (type)	Output Shape	Param #
conv3d (Conv3D)	(None, 6, 60, 60, 64)	640
max_pooling3d (MaxPooling3D)	(None, 3, 30, 30, 64)	0
conv3d_1 (Conv3D)	(None, 3, 28, 28, 64)	36928
max_pooling3d_1 (MaxPooling3	(None, 1, 14, 14, 64)	0
conv3d_2 (Conv3D)	(None, 1, 12, 12, 64)	36928
flatten (Flatten)	(None, 9216)	0
dropout (Dropout)	(None, 9216)	0
dense (Dense)	(None, 32)	294944
dense_1 (Dense)	(None, 32)	1056
dense_2 (Dense)	(None, 1)	33

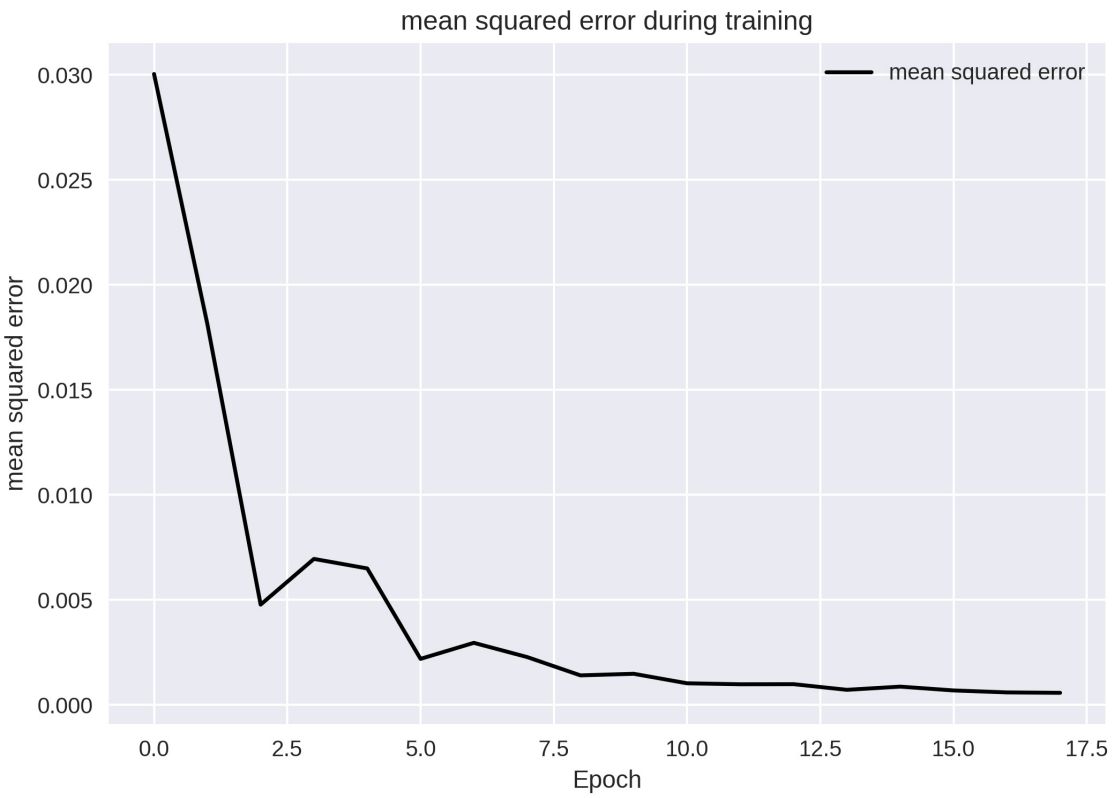
Total params: 370,529
Trainable params: 370,529
Non-trainable params: 0

Results

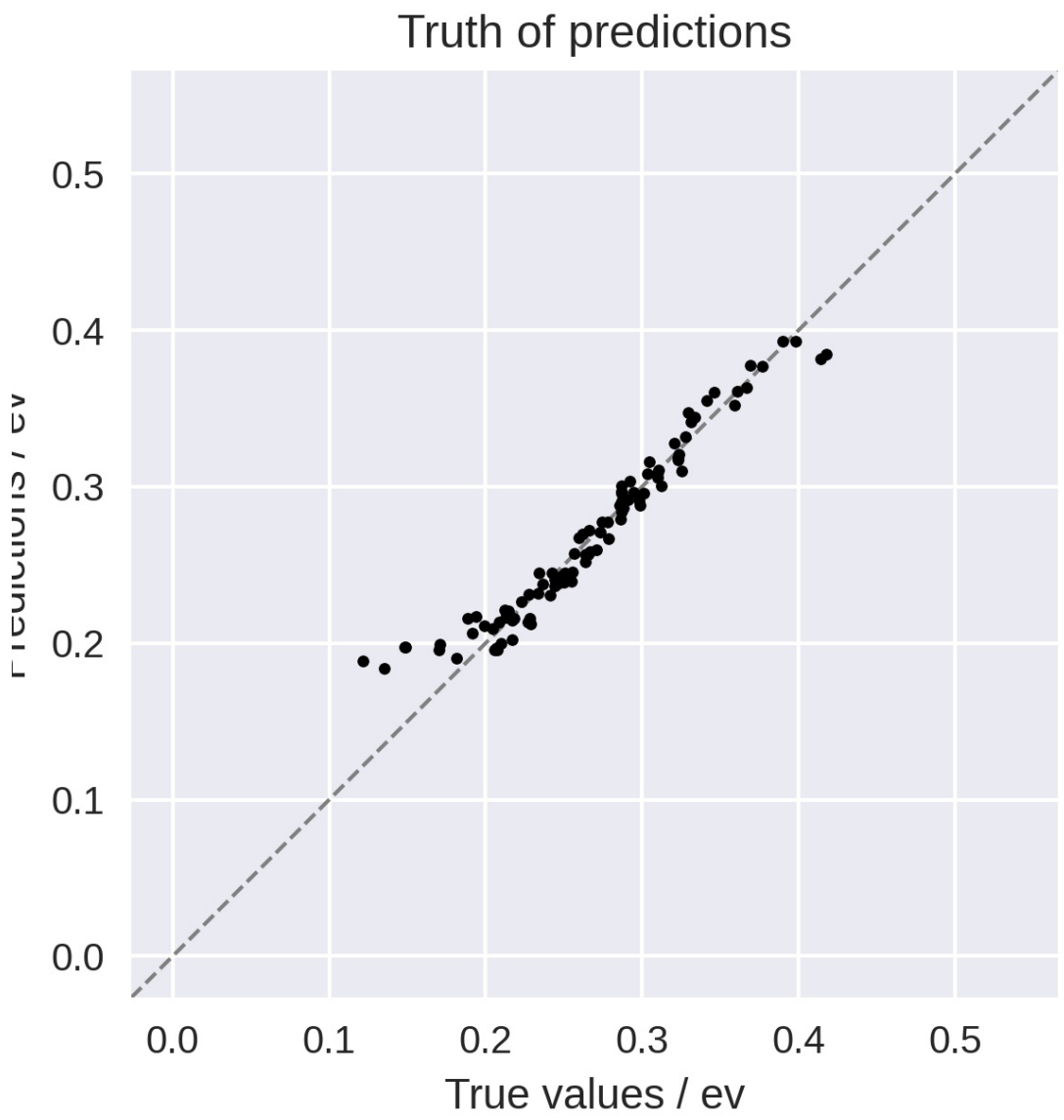
Training evaluation





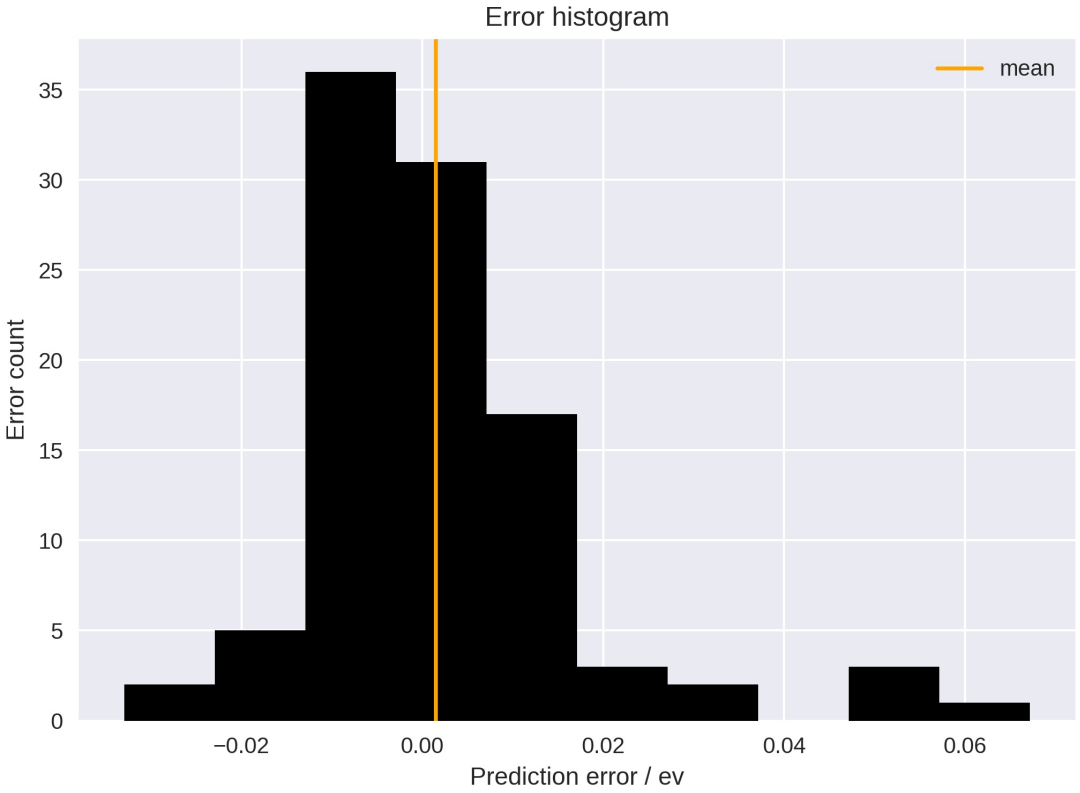


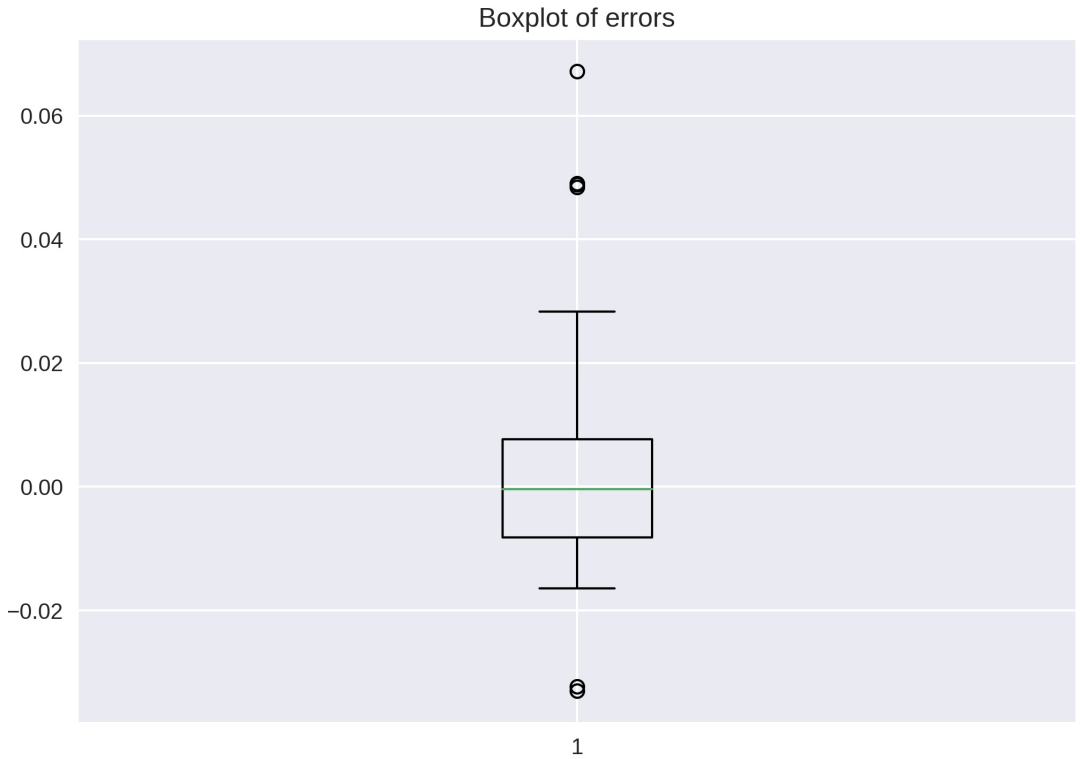
Model evaluation



mean absolute error: 0.0104
mean squared error: 0.0002

Error evaluation





Mean: 0.0015
Median: -0.0004
Minimum error: -0.033
Maximum error: 0.0672
Skewness: 1.5705
Kurtosis: 4.4751
Standard deviation: 0.0152
90% Confidence interval: [-0.001;0.004]

Log

```
2021-02-04 15:55:12,512:INFO:Starting model 3D_CNN_0.1.2_on_02-04-2021_15.55.12
2021-02-04 15:55:12,512:INFO:==================== Model info ====================
2021-02-04 15:55:12,512:INFO:Author: Michiel Jacobs
2021-02-04 15:55:12,512:INFO:Version: 0.1.2
2021-02-04 15:55:12,512:INFO:Modeltype: 3D CNN
2021-02-04 15:55:12,512:INFO:Maximum heavy atoms: 20
2021-02-04 15:55:12,512:INFO:Feature: Coulomb Matrix
2021-02-04 15:55:12,512:INFO:Labels: Zero point energy
2021-02-04 15:55:12,512:INFO:DEVELOPMENT: True
2021-02-04 15:55:12,512:INFO:============== Step 1: loading data ==============
2021-02-04 15:55:19,683:INFO:Data loaded
2021-02-04 15:55:19,683:INFO:============== Step 2: data preprocessing ==============
2021-02-04 15:55:19,684:INFO:Trimming dataset...
2021-02-04 15:55:20,016:INFO:Loading arrays...
2021-02-04 15:55:20,048:INFO:Shuffeling data...
2021-02-04 15:55:20,049:INFO:Calculating maximum size of molecules...
2021-02-04 15:55:20,049:INFO:The maximumsize of molecules is 62
2021-02-04 15:55:20,049:INFO:Normalizing data...
2021-02-04 15:55:20,077:INFO:Tensorisation of the coulomb matrices...
2021-02-04 15:55:34,047:INFO:Building channels...
2021-02-04 15:55:34,050:INFO:Calculating train test split...
2021-02-04 15:55:34,050:INFO:There are 500 entries in this dataset.
2021-02-04 15:55:34,050:INFO:Split ratio set to 0.8.
2021-02-04 15:55:34,050:INFO:Trainingset contains 400 molecules.
2021-02-04 15:55:34,050:INFO:Building train and test sets...
2021-02-04 15:55:34,051:INFO:Converting train features to tf.tensors...
2021-02-04 16:25:54,759:INFO:Converting test features to tf.tensors...
2021-02-04 16:33:30,667:INFO:Converting train labels to array...
2021-02-04 16:33:30,668:INFO:Converting test labels to array...
2021-02-04 16:33:30,668:INFO:============== Step 3: Model compilation ==============
2021-02-04 16:33:30,668:INFO:Building model...
2021-02-04 16:33:31,227:INFO:Model: "sequential"
2021-02-04
16:33:31,227:INFO:_________________________________________________________________
2021-02-04 16:33:31,227:INFO:Layer (type)                 Output Shape              Param #
2021-02-04
16:33:31,228:INFO:=================================================================
==========
2021-02-04 16:33:31,228:INFO:conv3d (Conv3D)              (None, 6, 60, 60, 64)     640
2021-02-04
16:33:31,228:INFO:_________________________________________________________________
2021-02-04 16:33:31,228:INFO:max_pooling3d (MaxPooling3D) (None, 3, 30, 30, 64)     0
2021-02-04
16:33:31,228:INFO:_________________________________________________________________
2021-02-04 16:33:31,228:INFO:conv3d_1 (Conv3D)            (None, 3, 28, 28, 64)     36928
2021-02-04
16:33:31,228:INFO:_________________________________________________________________
2021-02-04 16:33:31,228:INFO:max_pooling3d_1 (MaxPooling3 (None, 1, 14, 14, 64)     0
2021-02-04
```

16:33:31,228:INFO:_____
2021-02-04 16:33:31,228:INFO:conv3d_2 (Conv3D) (None, 1, 12, 12, 64) 36928
2021-02-04
16:33:31,229:INFO:_____
2021-02-04 16:33:31,229:INFO:flatten (Flatten) (None, 9216) 0
2021-02-04
16:33:31,229:INFO:_____
2021-02-04 16:33:31,229:INFO:dropout (Dropout) (None, 9216) 0
2021-02-04
16:33:31,229:INFO:_____
2021-02-04 16:33:31,229:INFO:dense (Dense) (None, 32) 294944
2021-02-04
16:33:31,229:INFO:_____
2021-02-04 16:33:31,229:INFO:dense_1 (Dense) (None, 32) 1056
2021-02-04
16:33:31,229:INFO:_____
2021-02-04 16:33:31,229:INFO:dense_2 (Dense) (None, 1) 33
2021-02-04
16:33:31,231:INFO:===
==========
2021-02-04 16:33:31,231:INFO:Total params: 370,529
2021-02-04 16:33:31,232:INFO:Trainable params: 370,529
2021-02-04 16:33:31,232:INFO:Non-trainable params: 0
2021-02-04
16:33:31,232:INFO:_____
2021-02-04 16:33:31,232:INFO:Compiling the model...
2021-02-04 16:33:31,238:INFO:============== Step 4: Model training ==============
2021-02-04 16:33:31,238:INFO:Enabeling early stopping...
2021-02-04 16:33:31,238:INFO:Start training...
2021-02-04 16:34:11,050:INFO:Plotting loss...
2021-02-04 16:34:11,449:INFO:Plotting metric mean_absolute_error
2021-02-04 16:34:11,641:INFO:Plotting metric mean_squared_error
2021-02-04 16:34:11,834:INFO:============== Step 5: Model evaluation ==============
2021-02-04 16:34:11,834:INFO:Evaluating model...
2021-02-04 16:34:12,087:INFO:Test scores:
2021-02-04 16:34:12,088:INFO:{'loss': 0.00023061165120452642,
'mean_absolute_error': 0.010409644804894924,
'mean_squared_error': 0.00023061165120452642}
2021-02-04 16:34:12,088:INFO:Making test predictions...
2021-02-04 16:34:12,564:INFO:Plotting ToP plot...
2021-02-04 16:34:12,796:INFO:Plotting Error histogram plot...
2021-02-04 16:34:13,002:INFO:Plotting boxplot...
2021-02-04 16:34:13,158:INFO:============== Step 6: Saving, reporting and cleanup
==============
2021-02-04 16:34:13,158:INFO:Saving model...
2021-02-04 16:34:13,537:WARNING:From
/apps/brussel/CO7/broadwell/software/TensorFlow/2.3.1-foss-2020a-Python-3.8.2/lib/python3.8/site-packages/tensorflow/python/training/tracking/tracking.py:111: Model.state_updates (from
tensorflow.python.keras.engine.training) is deprecated and will be removed in a future version.
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.

2021-02-04 16:34:13,542:WARNING:From /apps/brussel/CO7/broadwell/software/TensorFlow/2.3.1-foss-2020a-Python-3.8.2/lib/python3.8/site-packages/tensorflow/python/training/tracking/tracking.py:111: Layer.updates (from tensorflow.python.keras.engine.base_layer) is deprecated and will be removed in a future version.
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
2021-02-04 16:34:14,793:INFO:Assets written to: /scratch/brussel/102/vsc10255/Experimental-Reactivity-Prediction/code/models/../../models/CM_ZPE_3DCNN/3D_CNN_0.1.2_on_02-04-2021_15.55.12.tf/assets
2021-02-04 16:34:15,056:INFO:Model saved.
2021-02-04 16:34:15,057:INFO:Generating report...
2021-02-04 16:34:17,469:INFO:Page break on page 5 at y=271 for element of height 116