paper06_creating_models_for_train

February 27, 2022

1 Timeseries Testing and Modeling

Jorge III Altamirano-Astorga, Ita-Andehui Santiago, Luz Aurora Hernández.

Prof.: Edgar Francisco Román-Rangel.

```
                     temperature    pressure  …  wind_speed    wind_deg
datetime                                      …
2021-02-12 06:00:00    21.530000  777.410000  …    2.565310  109.799270
2021-02-12 06:05:00    21.689773  777.389432  …    2.456273  105.132299

[2 rows x 20 columns]
```

1.1 Timeseries

We use the the `timeseries_dataset_from_array` function from Keras Timeseries modeling functions. This function creates dataframes with sliding windows over time as an array.

This function work as follow:

```
tf.keras.utils.timeseries_dataset_from_array(
    X, Y,              # our dataset
    sequence_length,   #  Length of the output sequences (in number of timesteps): we need to text different s
    sequence_stride=1, # 1 is the default value for s, data[i], data[i + s], data[i + 2 * s]
    sampling_rate=1,   # 1 is the default value for timesteps data[i], data[i + r], ... data[i + sequence_leng
    shuffle=False,     # shuffle the records sorting, we set it in the default false value as the original ord
    seed=None,         # we set a fixed seed to have repeatable experiments
    #...
)
```

On our research it is relevant to have an adequate sequence length. Then we can focus a brief research on empiric good times. Our proposal are:

1.1.1 1 Minute Resampling

- 2 days before: i.e. on our resampling for every 5 min we'd have 2880 records. This is because 2 $days \times 24$ $hours \times 60$ $min \div 1$ min
- 7 days before: i.e. on our resampling for every 5 min we'd have 10,080 records. This is because 7 $days \times 24$ $hours \times 60$ $min \div 1$ min
- 15 days before: i.e. on our resampling for every 5 min we'd have 21600 records. This is because 15 $days \times 24$ $hours \times 60$ $min \div 1$ min

1.1.2 2 Minute Resampling

- 2 days before: i.e. on our resampling for every 5 min we'd have 1,440 records. This is because 2 $days \times 24$ $hours \times 60$ $min \div 2$ min
- 7 days before: i.e. on our resampling for every 5 min we'd have 5040 records. This is because 7 $days \times 24$ $hours \times 60$ $min \div 2$ min
- 15 days before: i.e. on our resampling for every 5 min we'd have 10,080 records. This is because 15 $days \times 24$ $hours \times 60$ $min \div 2$ min

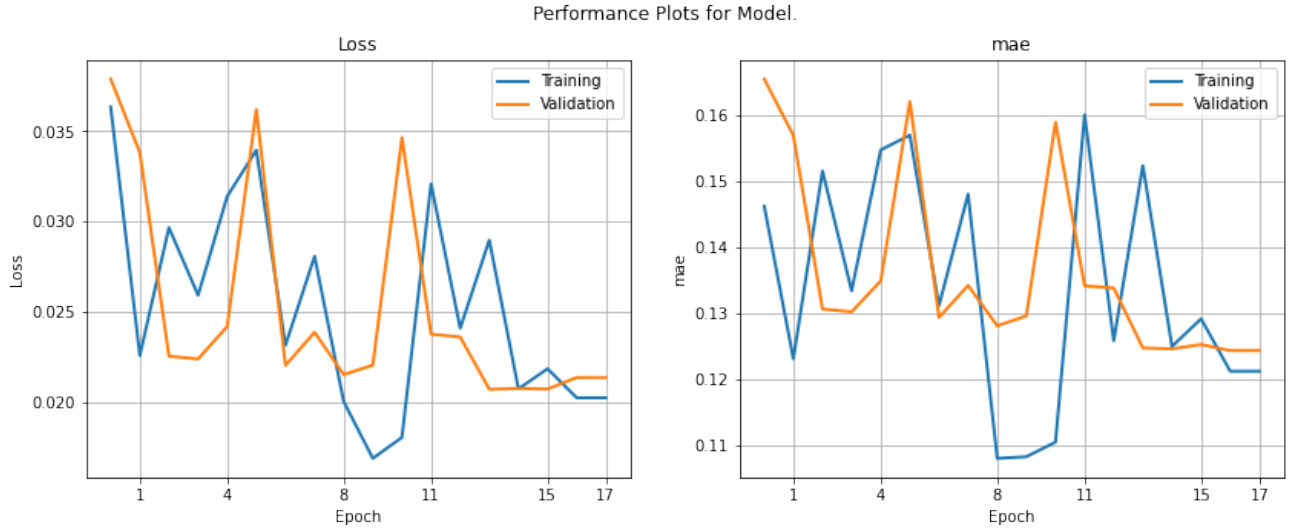
1.1.3 5 Minute Resampling

- 2 days before: i.e. on our resampling for every 5 min we'd have 576 records. This is because 2 $days \times 24\ hours \times 60\ min \div 5\ min$
- 7 days before: i.e. on our resampling for every 5 min we'd have 2016 records. This is because 7 $days \times 24\ hours \times 60\ min \div 5\ min$
- 15 days before: i.e. on our resampling for every 5 min we'd have 4320 records. This is because 15 $days \times 24\ hours \times 60\ min \div 5\ min$

We set this the number of days in a variable we set as `WINDOW_SIZE_DAYS`

1.2 5 Minute Resampling and 7 Days of History.

`Processing Time: 142.85 seconds.`

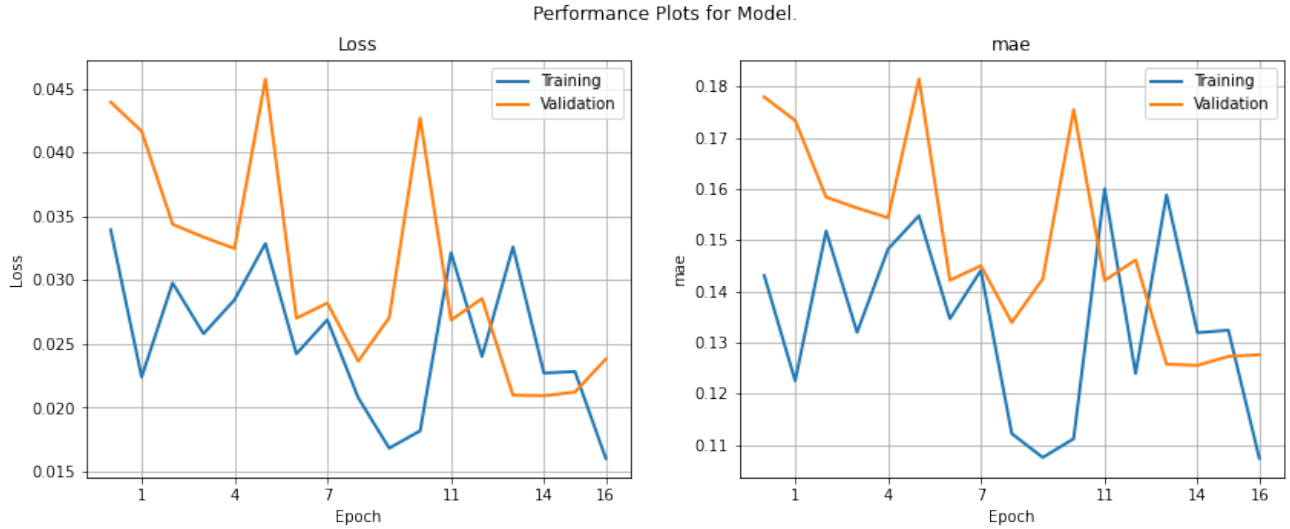


- Mean Absolute Error in Real Scale of the last Epoch: **61.6208 IAQ points**.

1.3 5 Minute Resampling and 2 Days of History with 2 Sampling Rate.

On our previous examples this quickly becomes unmanageable the we propose skipping some records and getting the previous hour by setting the `sampling_rate` parameter of 10 minute.

`Processing Time: 125.48 seconds.`

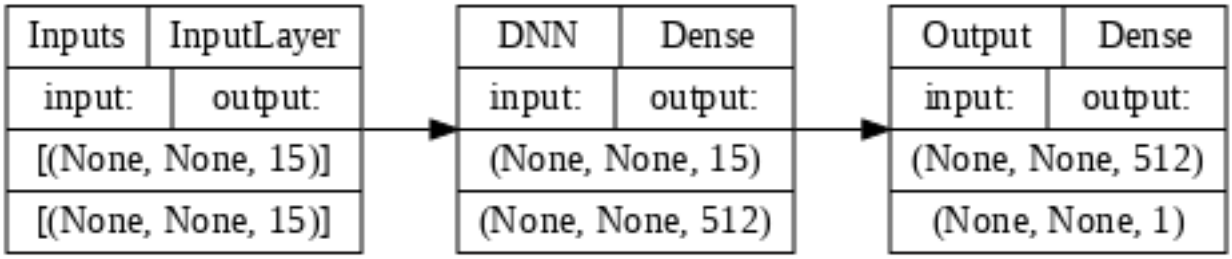


- Mean Absolute Error in Real Scale of the last Epoch: **63.2240 IAQ points**.

1.4 Model DNN01

Time series parameters:

```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```

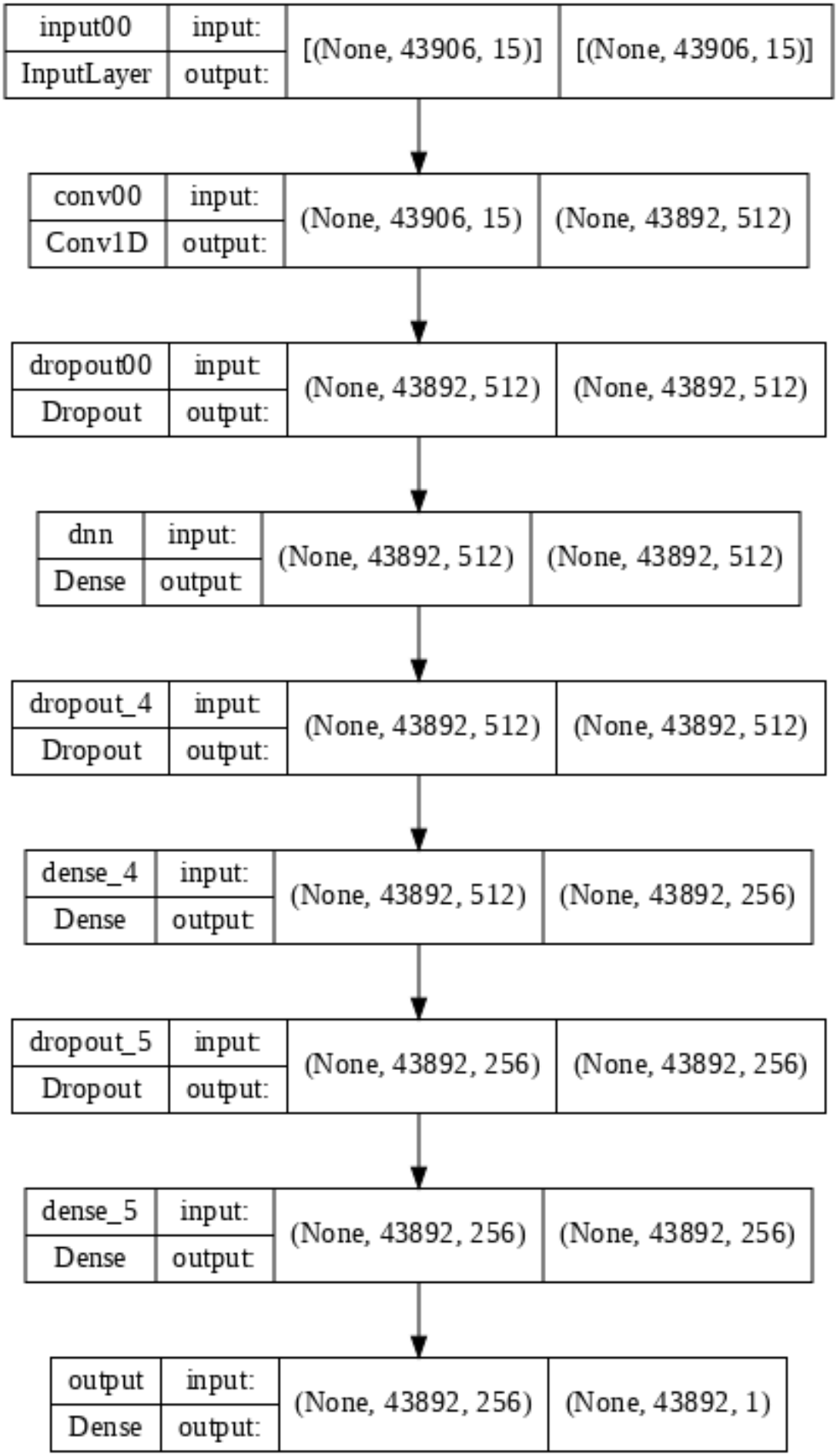



```
Saving file: drive/MyDrive/Colab Notebooks/proyecto-final/models-paper/scaler-
iaq.dill… Done!
```

1.5 Model Best 03a

Time series parameters:

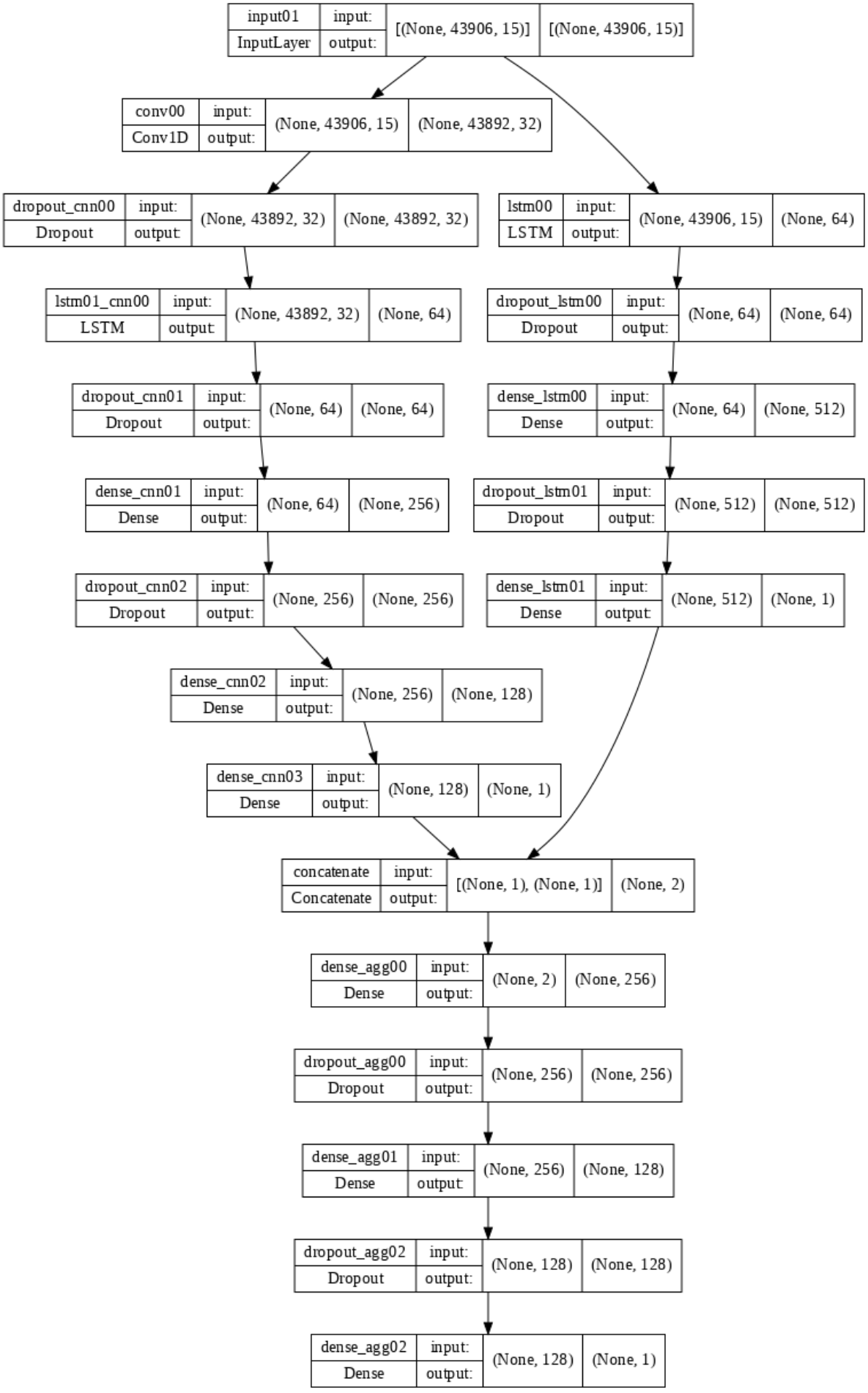
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```


input00
InputLayer
input:
output:
[(None, 43906, 15)]
[(None, 43906, 15)]
conv00
Conv1D
input:
output:
(None, 43906, 15)
(None, 43892, 512)
dropout00
Dropout
input:
output:
(None, 43892, 512)
(None, 43892, 512)
dnn
Dense
input:
output:
(None, 43892, 512)
(None, 43892, 512)
dropout_4
Dropout
input:
output:
(None, 43892, 512)
(None, 43892, 512)
dense_4
Dense
input:
output:
(None, 43892, 512)
(None, 43892, 256)
dropout_5
Dropout
input:
output:
(None, 43892, 256)
(None, 43892, 256)
dense_5
Dense
input:
output:
(None, 43892, 256)
(None, 43892, 256)
output
Dense
input:
output:
(None, 43892, 256)
(None, 43892, 1)

1.6 Model Best 03b

Time series parameters:

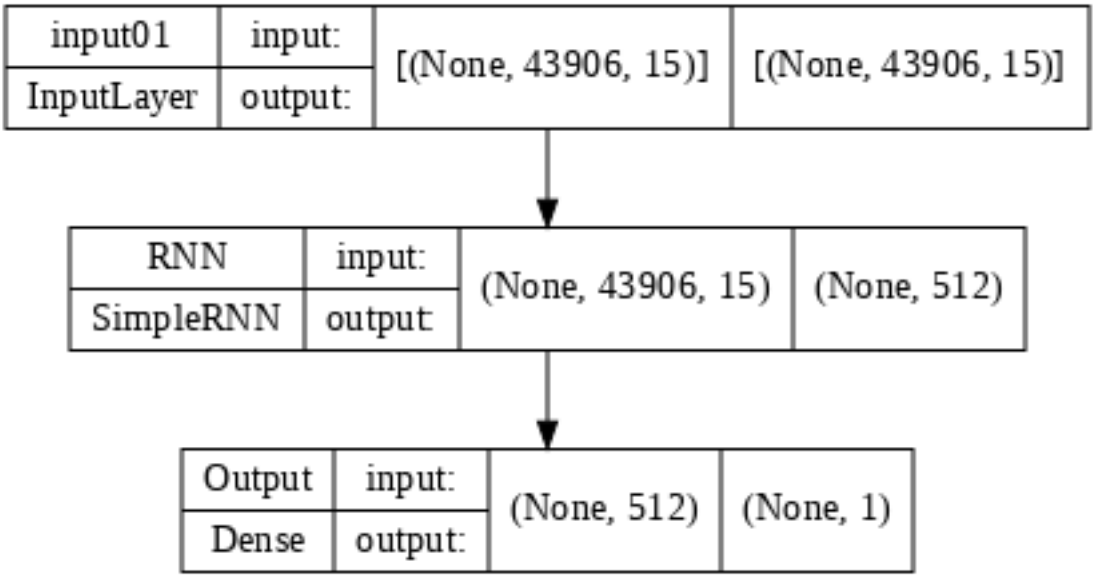
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```


input01 | input: | [(None, 43906, 15)] | [(None, 43906, 15)]
InputLayer | output:
conv00 | input: | (None, 43906, 15) | (None, 43892, 32)
Conv1D | output:
dropout_cnn00 | input: | (None, 43892, 32) | (None, 43892, 32)
Dropout | output:
lstm00 | input: | (None, 43906, 15) | (None, 64)
LSTM | output:
lstm01_cnn00 | input: | (None, 43892, 32) | (None, 64)
LSTM | output:
dropout_lstm00 | input: | (None, 64) | (None, 64)
Dropout | output:
dropout_cnn01 | input: | (None, 64) | (None, 64)
Dropout | output:
dense_lstm00 | input: | (None, 64) | (None, 512)
Dense | output:
dense_cnn01 | input: | (None, 64) | (None, 256)
Dense | output:
dropout_lstm01 | input: | (None, 512) | (None, 512)
Dropout | output:
dropout_cnn02 | input: | (None, 256) | (None, 256)
Dropout | output:
dense_lstm01 | input: | (None, 512) | (None, 1)
Dense | output:
dense_cnn02 | input: | (None, 256) | (None, 128)
Dense | output:
dense_cnn03 | input: | (None, 128) | (None, 1)
Dense | output:
concatenate | input: | [(None, 1), (None, 1)] | (None, 2)
Concatenate | output:
dense_agg00 | input: | (None, 2) | (None, 256)
Dense | output:
dropout_agg00 | input: | (None, 256) | (None, 256)
Dropout | output:
dense_agg01 | input: | (None, 256) | (None, 128)
Dense | output:
dropout_agg02 | input: | (None, 128) | (None, 128)
Dropout | output:
dense_agg02 | input: | (None, 128) | (None, 1)
Dense | output:

1.7 Model RNN00

Time series parameters:

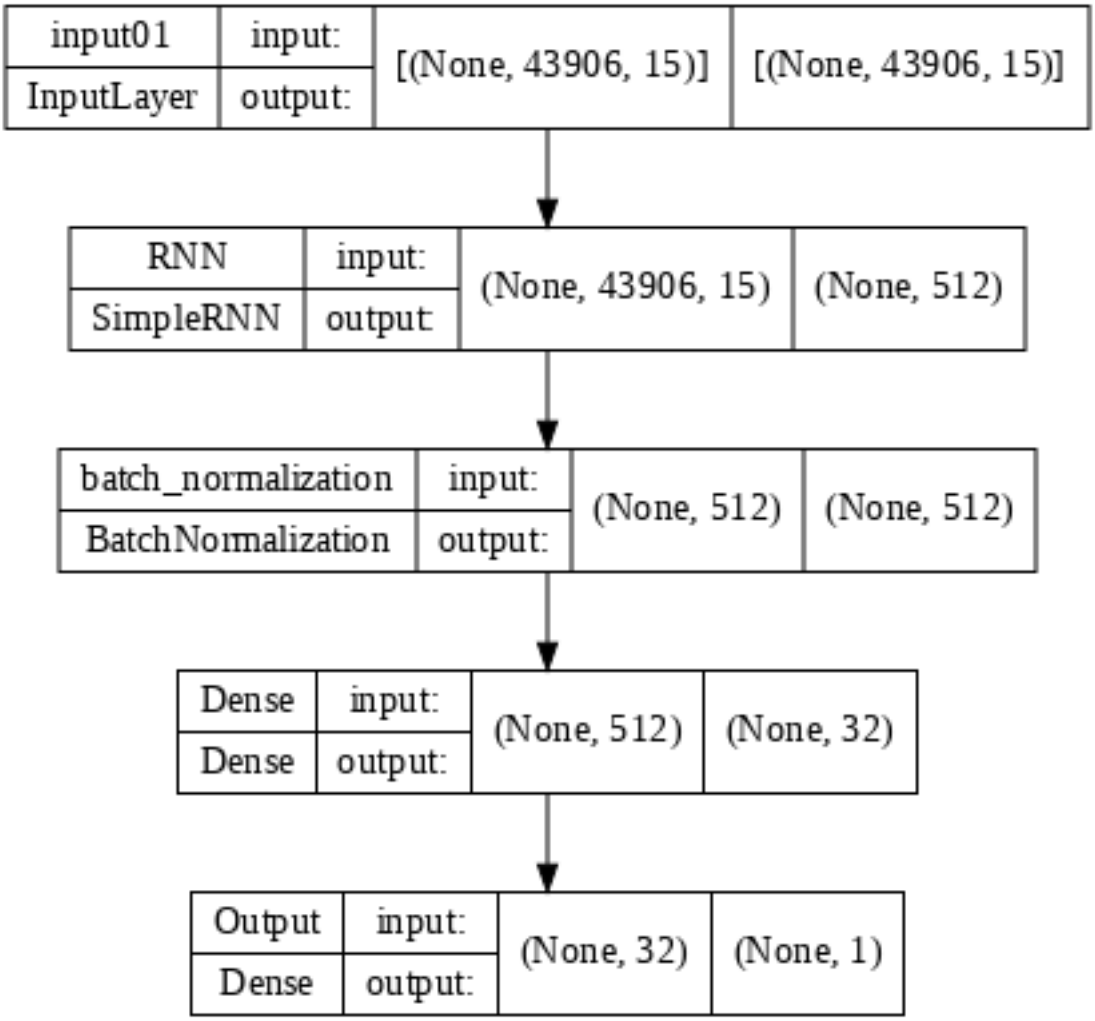
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.8 Model RNN02

Time series parameters:

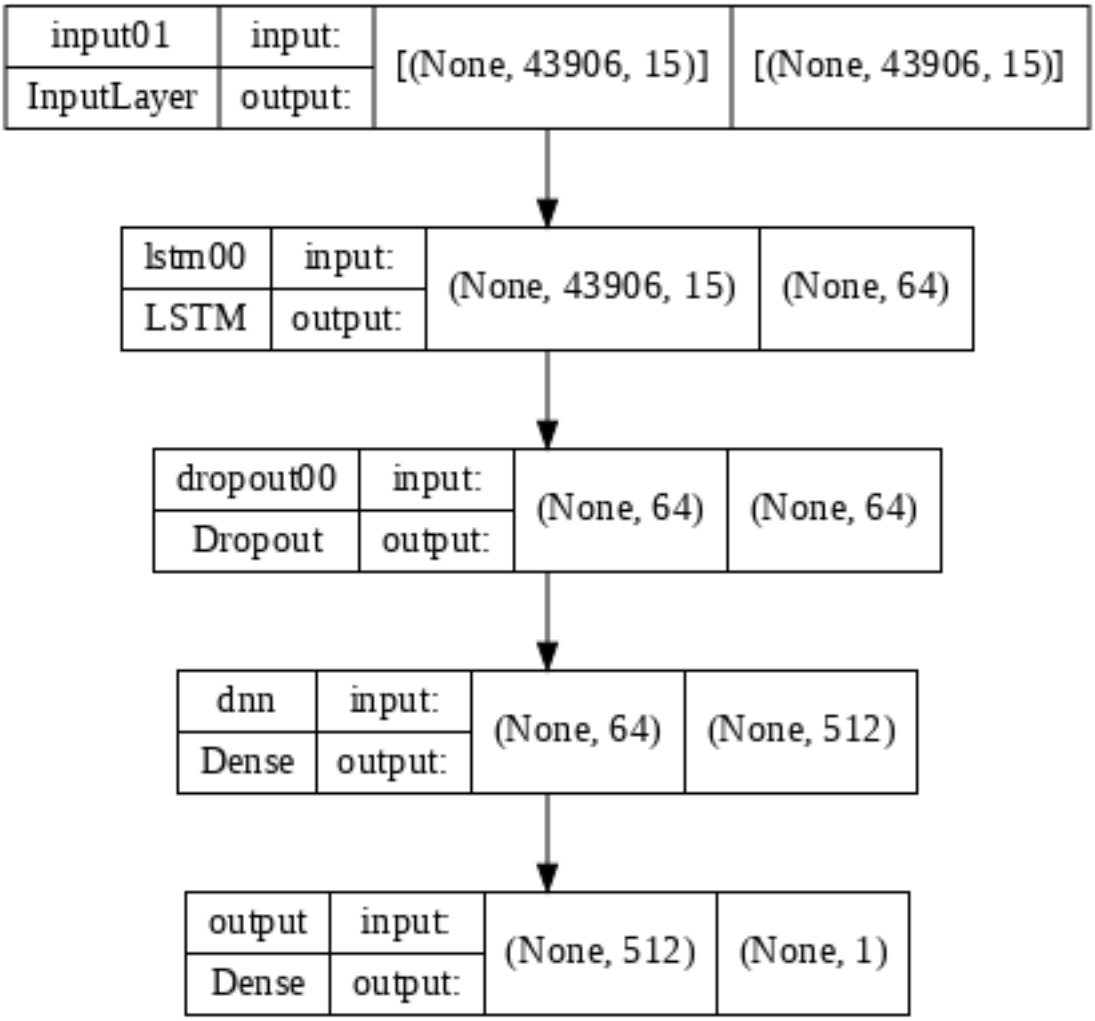
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.9 Model LSTM00

Time series parameters:

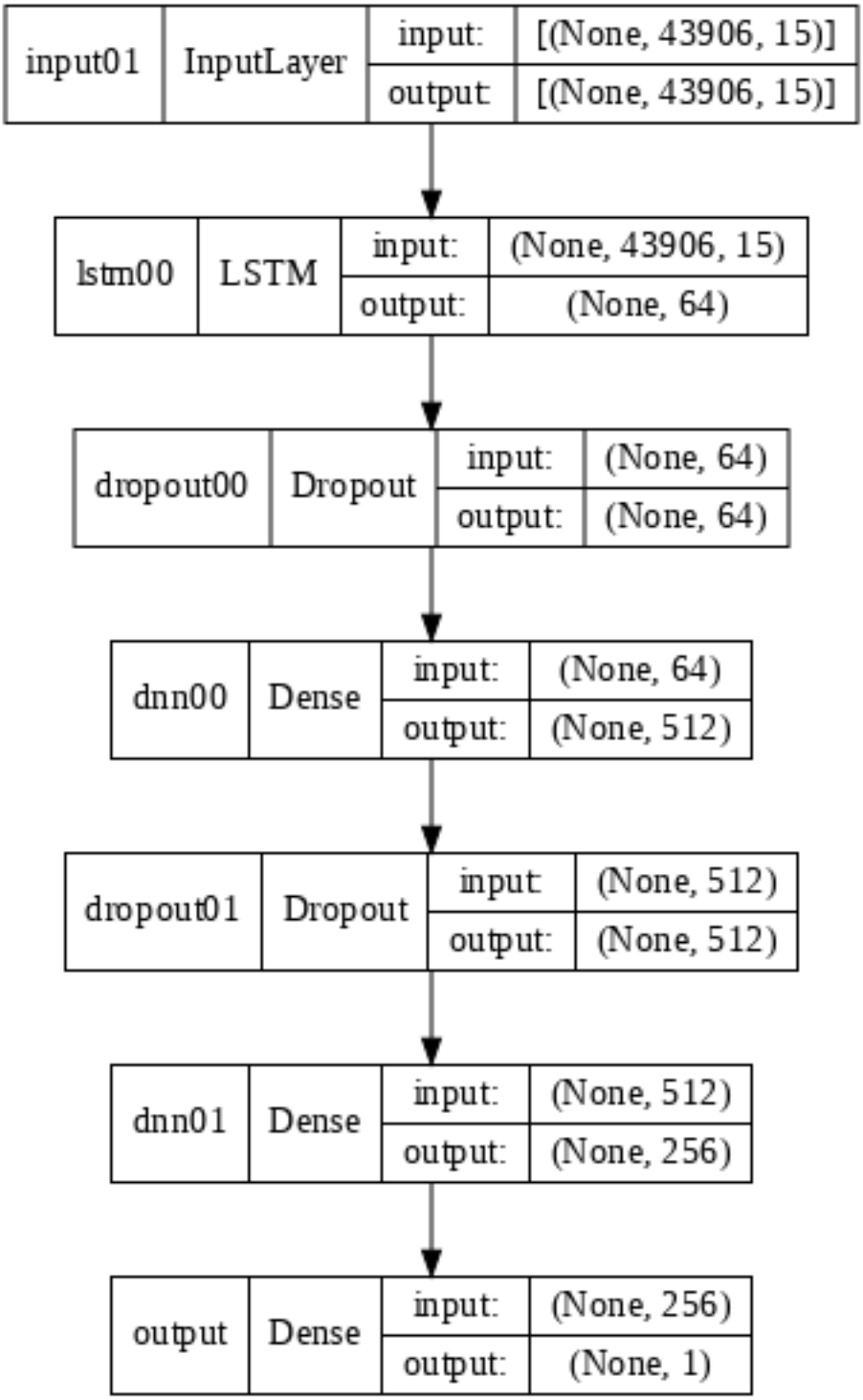
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.10 Model LSTM02

Time series parameters:

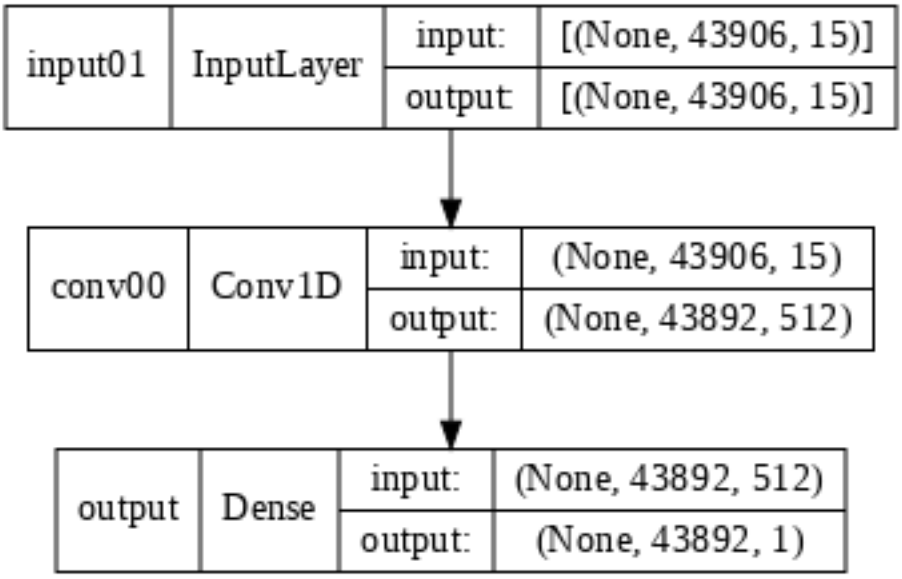
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.11 Model Conv00

Time series parameters:

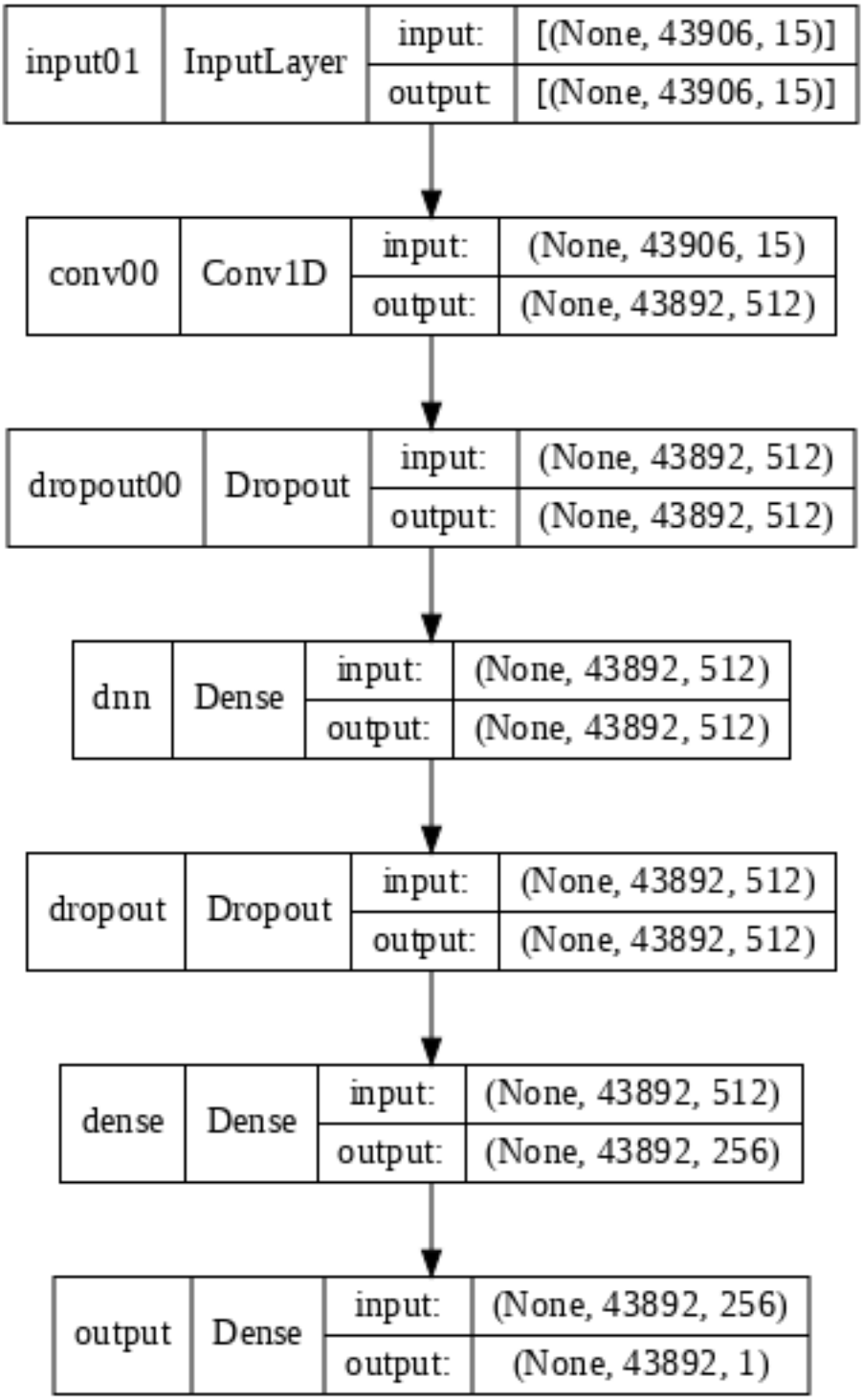
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.12 Model Conv02

Time series parameters:

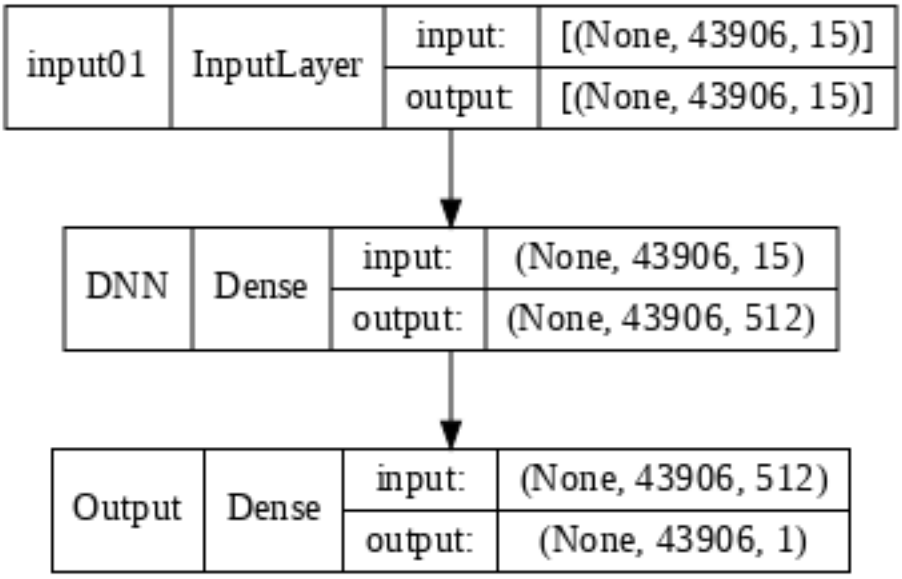
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.13 Model DNN00

Time series parameters:

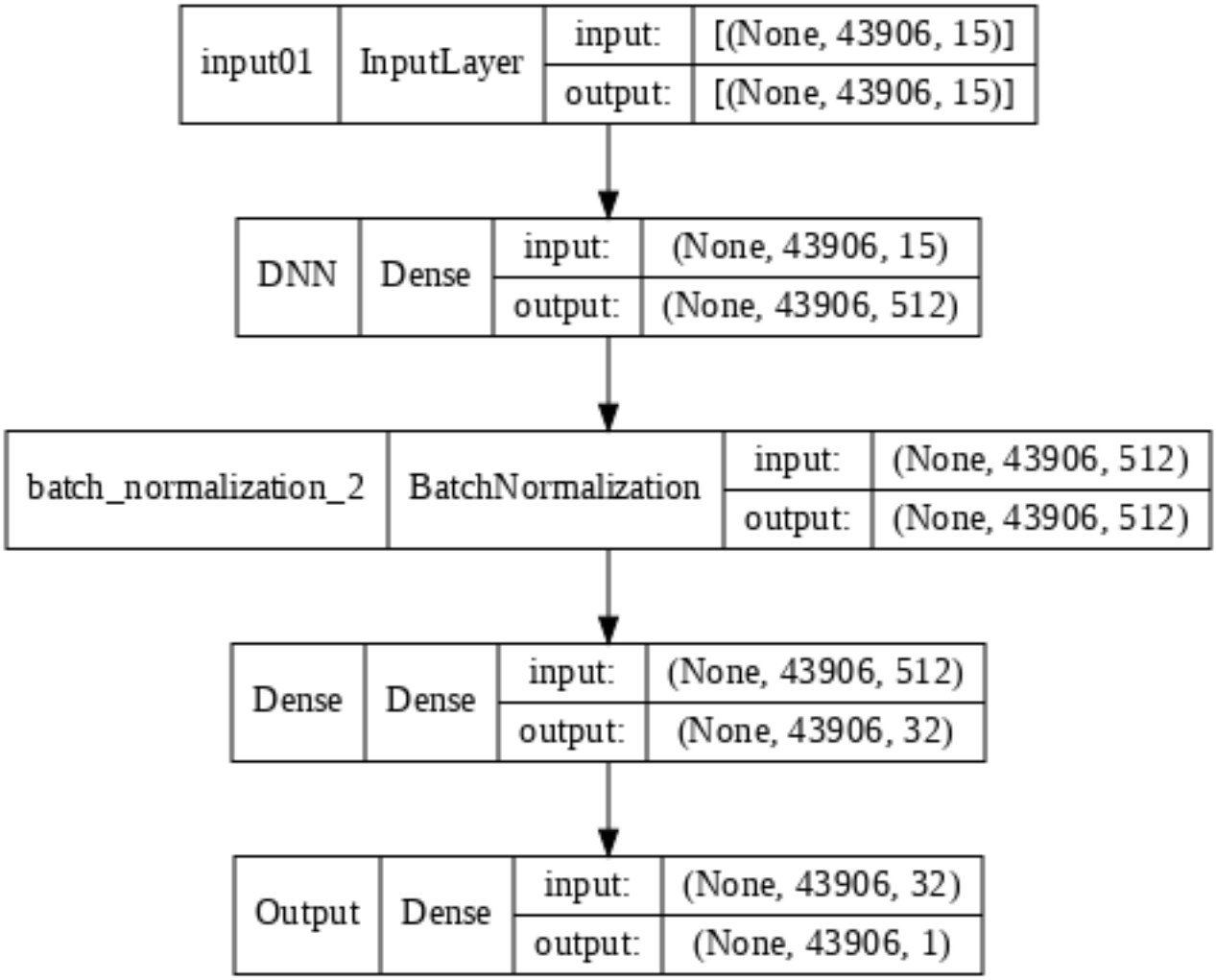
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.14 Model DNN02

Time series parameters:

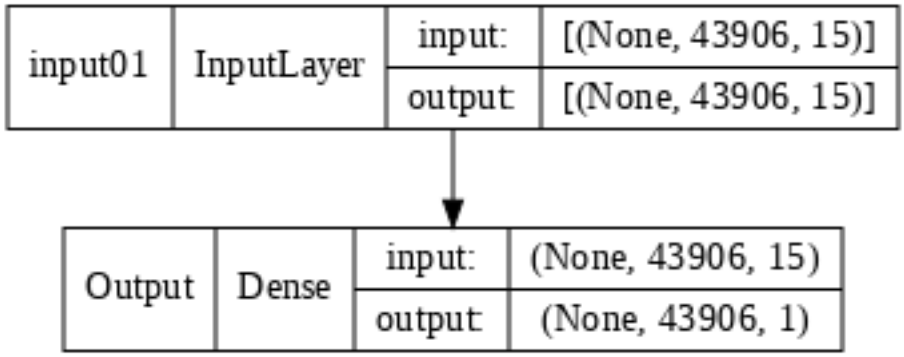
```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.15 Model Baseline

Time series parameters:

```
timeseries_dataset_from_array(  sequence_length=576,  sampling_rate=1,  batch_size=256,  seed=175904)
```



1.16 Fixing 6 Weeks Models

We found that the validation didn't run for some reason on our models when having a history of 6 weeks.

Reading the models:

```
Loading file model_file model_lstm00.h5… Done!
Loading file model_file model_conv00.h5… Done!
Loading file model_file model_dnn00.h5… Done!
Loading file model_file model_conv02.h5… Done!
Loading file model_file model_best01a.h5… Done!
```

We set the data source in order to get the:

- True y's.
- Get the validation x's to predict the $\hat{y}$.

```
6480
```

We proceed to get the predict the $\hat{y}$:

```
Predicting the y hats for model "model_lstm00"… Done!
Predicting the y hats for model "model_conv00"… Done!
Predicting the y hats for model "model_dnn00"… Done!
Predicting the y hats for model "model_conv02"… Done!
Predicting the y hats for model "model_best01a"… Done!
CPU times: user 22.7 s, sys: 4.25 s, total: 27 s
Wall time: 44.5 s
```

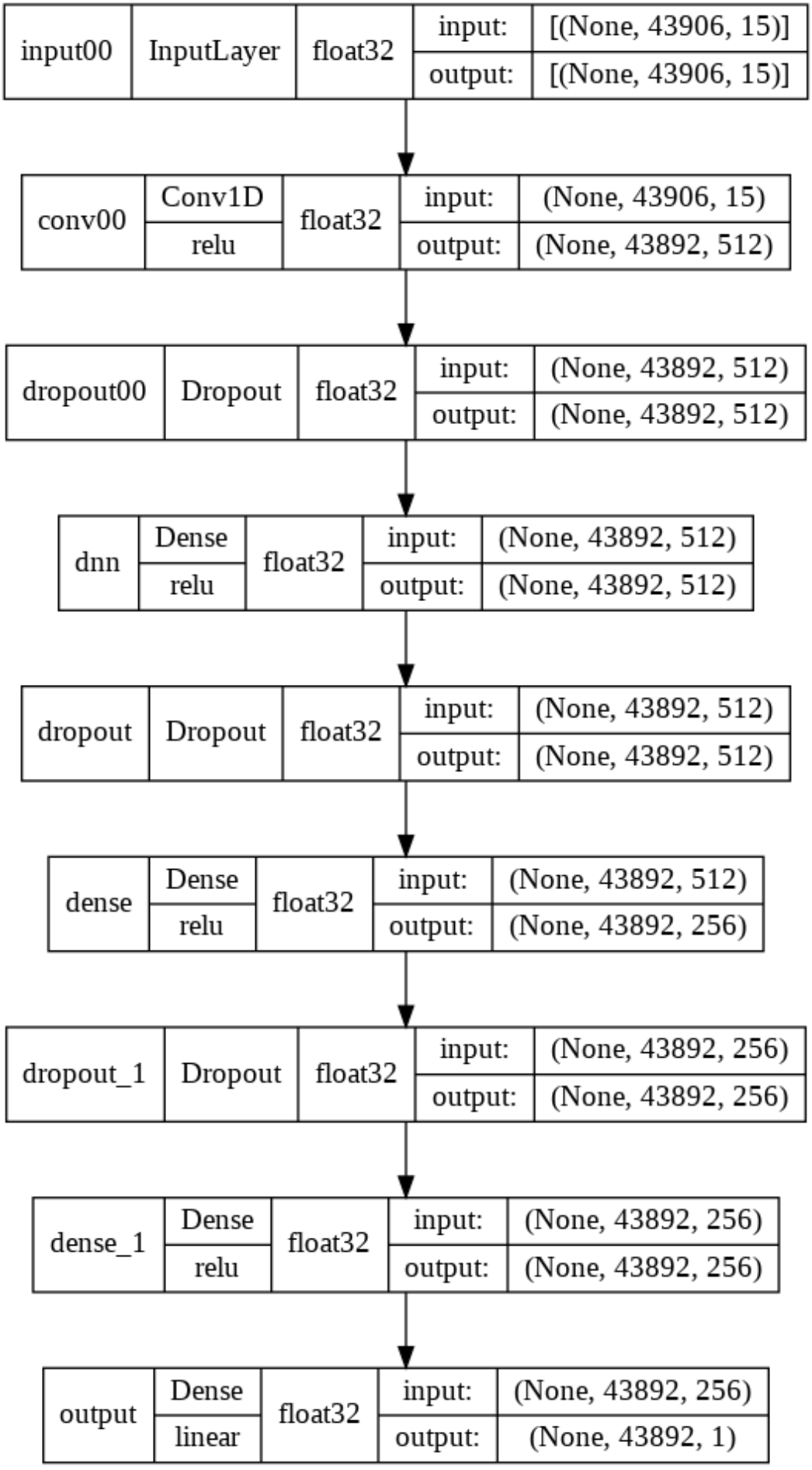
```
dict_keys(['model_lstm00', 'model_conv00', 'model_dnn00', 'model_conv02',
'model_best01a'])
```

```
2931
```

```
9410
```

```
array([[0.40914166],
       [0.40914166],
       [0.40914166],
```

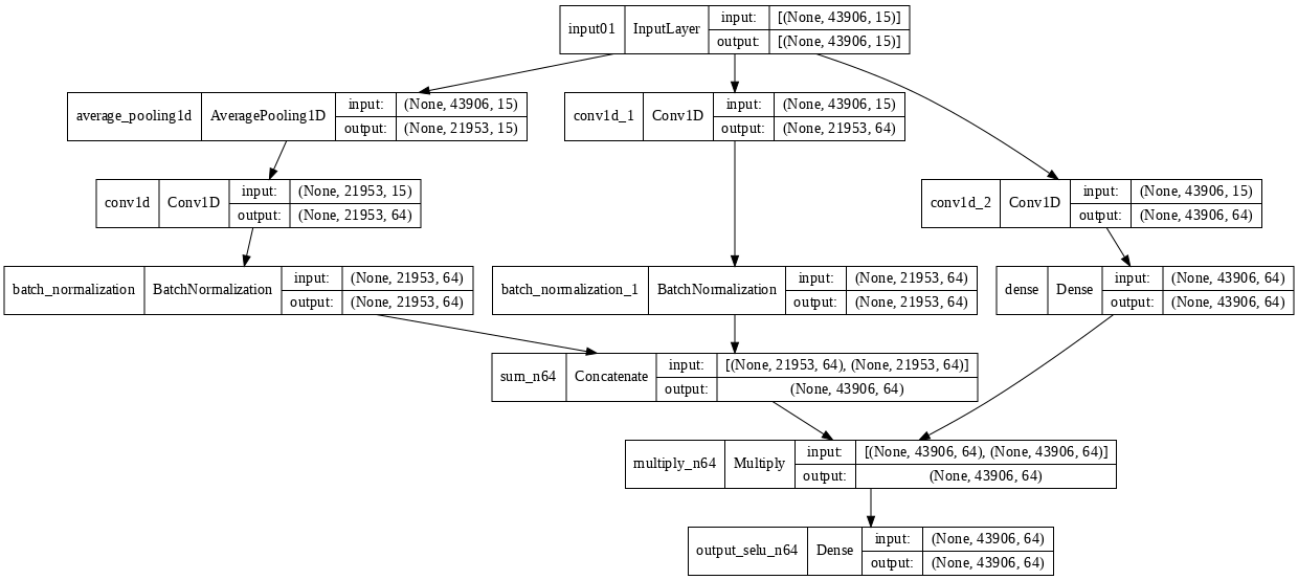
```
        …,
        [0.40914184],
        [0.40914187],
        [0.40914184]], dtype=float32)

 2931

 0.030596263103119924

Shape of model_lstm00: (2931, 1)
 Shape of model_lstm00: (2931,)
  mse=0.028633685371812847
  mae=0.14440499236337684
Shape of model_conv00: (2931, 6466, 1)
 Shape of model_conv00: (2931,)
  mse=0.10565433191409049
  mae=0.28295468567733634
Shape of model_dnn00: (2931, 6480, 1)
 Shape of model_dnn00: (2931,)
  mse=0.05798647852976501
  mae=0.2012153250123069
Shape of model_conv02: (2931, 6466, 1)
 Shape of model_conv02: (2931,)
  mse=0.026132470975462244
  mae=0.13538214853361374
Shape of model_best01a: (2931, 6466, 1)
 Shape of model_best01a: (2931,)
  mse=0.030596268615801434
  mae=0.14814169067075755

 0.030596268615801434

 (2931, 6466, 1)

 (2931,)
```


input00 InputLayer float32 input: [(None, 43906, 15)] output: [(None, 43906, 15)]
conv00 Conv1D relu float32 input: (None, 43906, 15) output: (None, 43892, 512)
dropout00 Dropout float32 input: (None, 43892, 512) output: (None, 43892, 512)
dnn Dense relu float32 input: (None, 43892, 512) output: (None, 43892, 512)
dropout Dropout float32 input: (None, 43892, 512) output: (None, 43892, 512)
dense Dense relu float32 input: (None, 43892, 512) output: (None, 43892, 256)
dropout_1 Dropout float32 input: (None, 43892, 256) output: (None, 43892, 256)
dense_1 Dense relu float32 input: (None, 43892, 256) output: (None, 43892, 256)
output Dense linear float32 input: (None, 43892, 256) output: (None, 43892, 1)

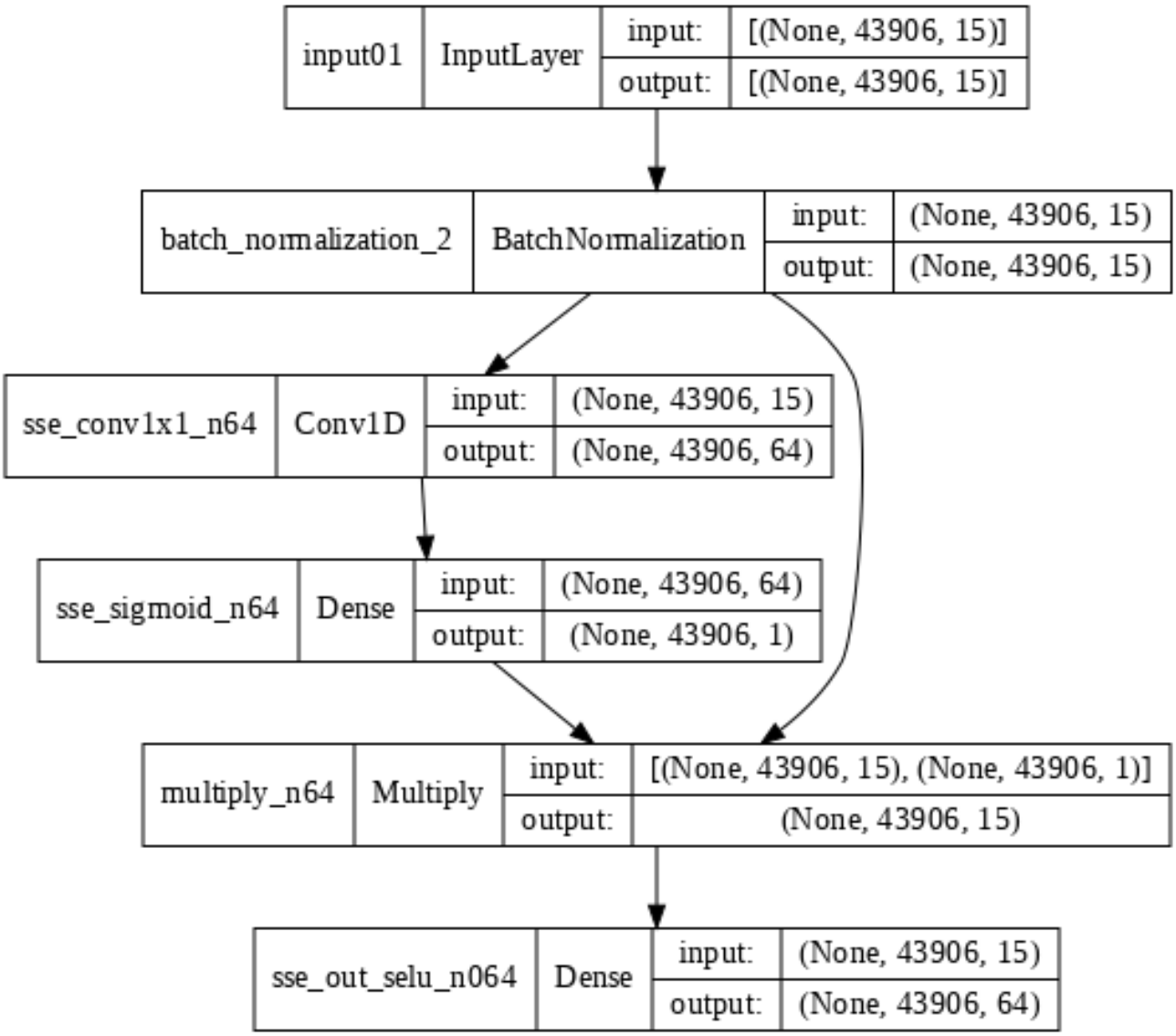
```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-106-73be1b5df45f> in <module>()
----> 1 models["model_best01a"]

TypeError: list indices must be integers or slices, not str
```

1.17 Model ParNet0

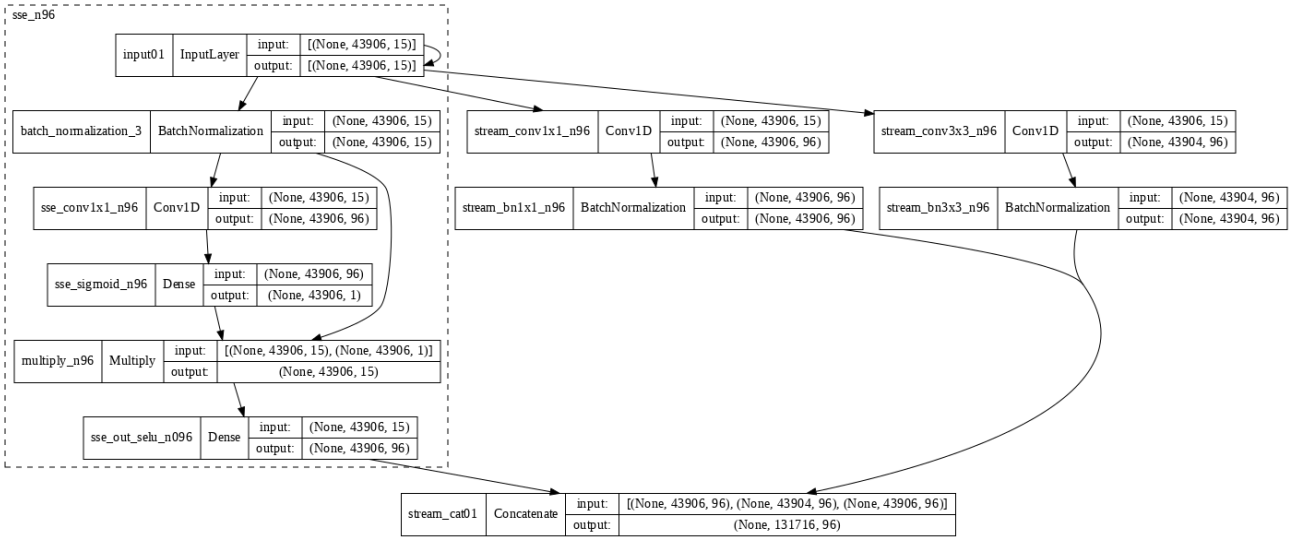
Downsampling



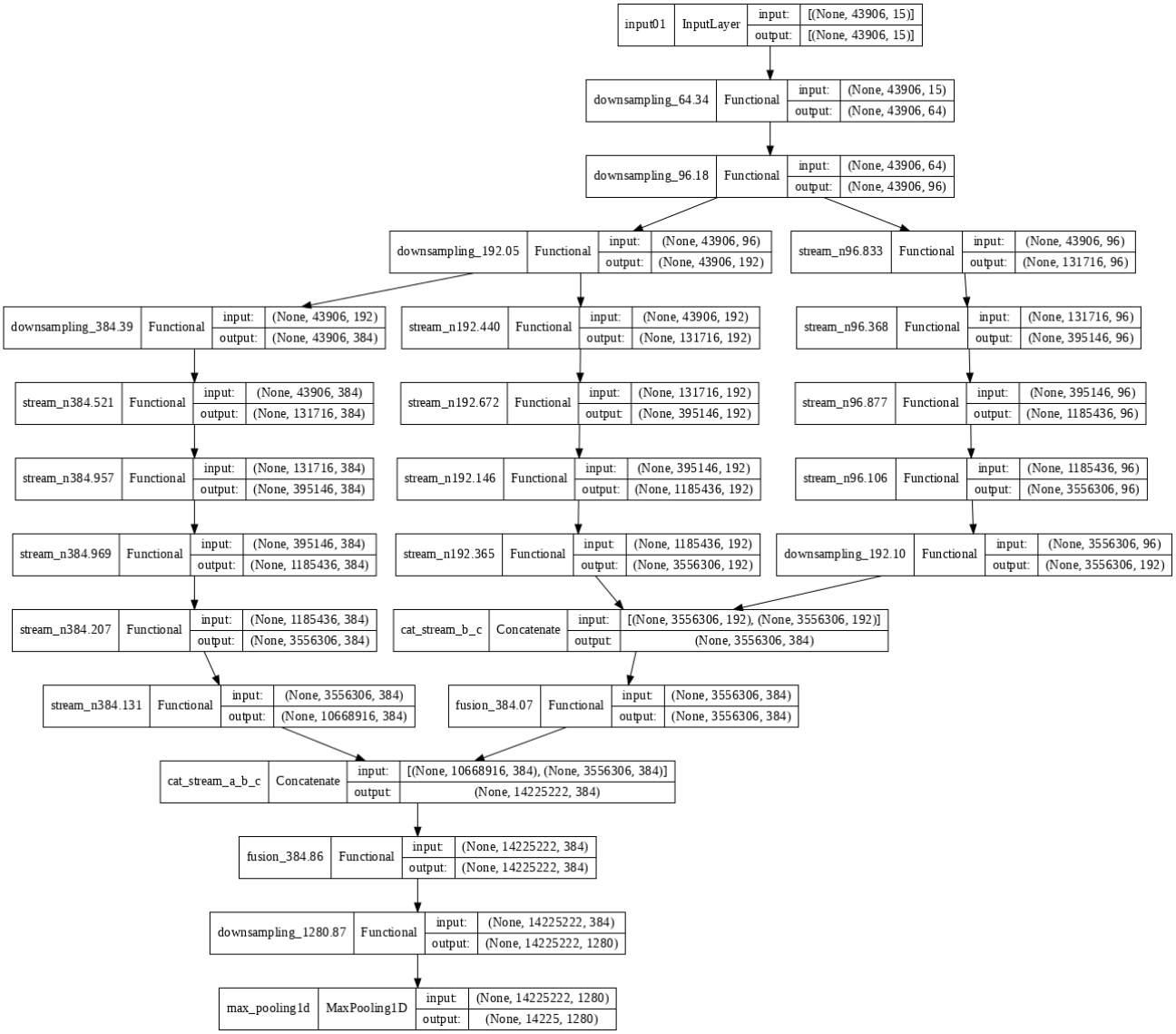
SSE


input01 InputLayer input: [(None, 43906, 15)] output: [(None, 43906, 15)]
batch_normalization_2 BatchNormalization input: (None, 43906, 15) output: (None, 43906, 15)
sse_conv1x1_n64 Conv1D input: (None, 43906, 15) output: (None, 43906, 64)
sse_sigmoid_n64 Dense input: (None, 43906, 64) output: (None, 43906, 1)
multiply_n64 Multiply input: [(None, 43906, 15), (None, 43906, 1)] output: (None, 43906, 15)
sse_out_selu_n064 Dense input: (None, 43906, 15) output: (None, 43906, 64)

Stream


sse_n96
input01 InputLayer input: [(None, 43906, 15)] output: [(None, 43906, 15)]
batch_normalization_3 BatchNormalization input: (None, 43906, 15) output: (None, 43906, 15)
stream_conv1x1_n96 Conv1D input: (None, 43906, 15) output: (None, 43906, 96)
stream_conv3x3_n96 Conv1D input: (None, 43906, 15) output: (None, 43904, 96)
sse_conv1x1_n96 Conv1D input: (None, 43906, 15) output: (None, 43906, 96)
stream_bn1x1_n96 BatchNormalization input: (None, 43906, 96) output: (None, 43906, 96)
stream_bn3x3_n96 BatchNormalization input: (None, 43904, 96) output: (None, 43904, 96)
sse_sigmoid_n96 Dense input: (None, 43906, 96) output: (None, 43906, 1)
multiply_n96 Multiply input: [(None, 43906, 15), (None, 43906, 1)] output: (None, 43906, 15)
sse_out_selu_n096 Dense input: (None, 43906, 15) output: (None, 43906, 96)
stream_cat01 Concatenate input: [(None, 43906, 96), (None, 43904, 96), (None, 43906, 96)] output: (None, 131716, 96)

Sequence




```
/usr/local/lib/python3.7/dist-packages/keras/engine/functional.py:1410:
CustomMaskWarning: Custom mask layers require a config and must override
get_config. When loading, the custom mask layer must be passed to the
custom_objects argument.
  layer_config = serialize_layer_fn(layer)
```

Job Training

256

0epoch [00:00, ?epoch/s]

0batch [00:00, ?batch/s]

Epoch 1/5

```
---------------------------------------------------------------------------
InvocationException                       Traceback (most recent call last)
<ipython-input-16-c6465fb85522> in <module>()
```

```
      3                               metrics=["mse", "mae"],
      4                               epochs=5, steps_per_epoch=2,
----> 5                               batch_size=BATCH_SIZE, base_dir=base_url)


<ipython-input-1-477fa0ed5883> in train_model(model, train_data, validation_data, epochs, batch_size,␣
 ↪steps_per_epoch, loss, optimizer, metrics, verbose, base_dir)
    119   display(plot_model(model, to_file=os.path.join(base_url,
    120                                                  f"data/{model.name}.png"),
--> 121            dpi=72, rankdir="TB", show_shapes=True, expand_nested=True))
    122   cbk = TqdmCallback()
    123   tiempo = time.time()

/usr/local/lib/python3.7/dist-packages/keras/utils/vis_utils.py in plot_model(model, to_file,␣
 ↪show_shapes, show_dtype, show_layer_names, rankdir, expand_nested, dpi, layer_range,␣
 ↪show_layer_activations)
    434     extension = extension[1:]
    435   # Save image to disk.
--> 436   dot.write(to_file, format=extension)
    437   # Return the image as a Jupyter Image object, to be displayed in-line.
    438   # Note that we cannot easily detect whether the code is running in a

/usr/local/lib/python3.7/dist-packages/pydot_ng/__init__.py in write(self, path, prog, format)
   1777
   1778             else:
-> 1779                 fobj.write(self.create(prog, format))
   1780         finally:
   1781             if close:

/usr/local/lib/python3.7/dist-packages/pydot_ng/__init__.py in create(self, prog, format)
   1889             raise InvocationException(
   1890                 'Program terminated with status: %d. stderr follows: %s' % (
-> 1891                     status, stderr_output))
   1892         elif stderr_output:
   1893             print(stderr_output)

InvocationException: Program terminated with status: 1. stderr follows: Error: /tmp/tmp2lzpb8hx:␣
 ↪syntax error in line 7 near '.94'
```