# Performance Evaluation of Storm Intensity prediction

## Dictionary

* Lookback[t]h – the model is revealed t timesteps before realtime as input. If lookback is 0 or not mentioned it means the model only receives one timestep as input.
* Forecast[t]h – the model will predict t timesteps forward.
* Realtime – the model will predict 0 timesteps forward.
* No\_intensity – the intensity parameter will be left out from the model inputs.

## Naive Approach

### Experiment Setup

* A linear model with no parameters
* Mark .
* Mark .
* Mark .
* Mark
* We will define the Naïve function to predict forecasted intensity:

### Results

|  |  |
| --- | --- |
| NAÏVE RUN | Mean Absolute Error |
| Forecast1h | 1.697 |
| Forecast2h | 3.18 |
| Lookback1h-Forecast1h | 1.065 |
| Lookback1h-Forecast2h | 2.432 |
| Lookback3h-Forecast1h | 1.35 |
| Lookback3h-Forecast2h | 2.875 |

### Conclusion

We have set up a benchmark for the rest of the experiment. These are the basic results we use to compare our experiments to.

## First Attempt at Training a Simple NN

### Experiment Setup

* Training dataset years: 1990-2000
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Trained 4 different simple NNs:
  + Forecast1h
  + Forecast2h
  + Realtime
  + Forecast1h, No\_Intensity

### Results

|  |  |
| --- | --- |
| Simple NN | Validation Mean Absolute Error |
| Forecast1h | 6.75 |
| Forecast2h | 6.55 |
| Realtime | 6.61 |
| Forecast1h No\_Intensity | 6.62 |

### Training Process

Train Loss RMSE:A screen shot of a graph

Description automatically generated

Validation Mean Absolute Error (MAE): A graph with colorful lines

Description automatically generated

### Conclusion

All versions reached an MAE of around 6.5-7 for the intensity parameter.

Forecast2h performed similarly to Forecast1h, sometimes better.  
Training loss over iterations was choppy, indicating suboptimal training, likely due to varying parameter scales.

We conclude that training does not converge. Next we will attempt to fix these problems by normalizing the input data.

## Normalized Data Simple NN Approach

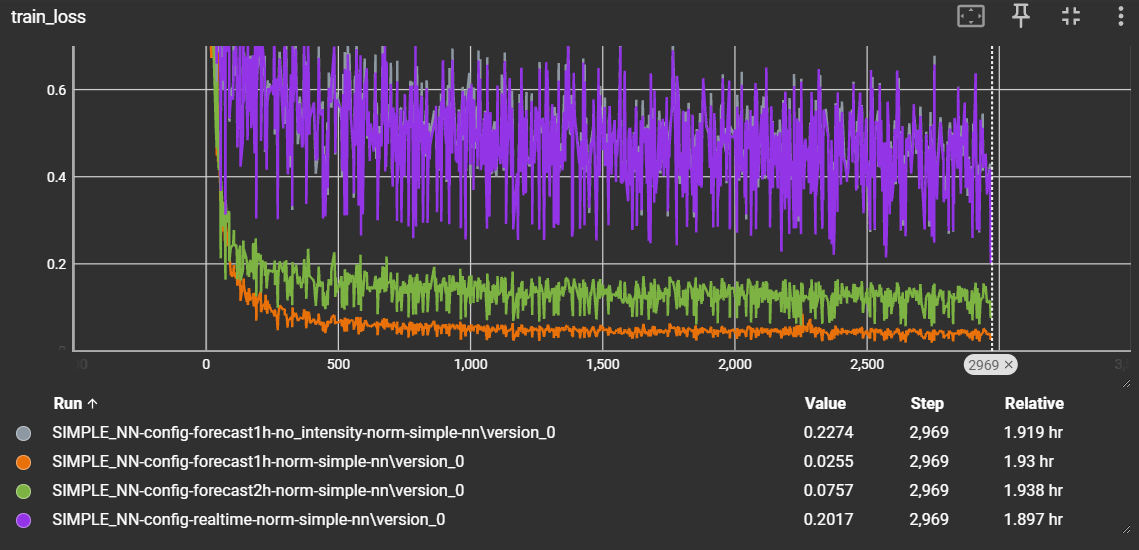
### Experiment Setup

* Training dataset years: 1990-2000
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Mean and STD of data is calculated across all training dataset for each parameter. Then we subtract the mean and divide by the STD so that each feature is normalized around mean=0 and std=1.
* Trained 4 different simple NNs on the normalized data:
  + Forecast1h
  + Forecast2h
  + Realtime
  + Forecast1h, No\_Intensity

### Results

|  |  |
| --- | --- |
| Simple NN Normalized Inputs | Validation Mean Absolute Error |
| Forecast1h | 1.74 |
| Forecast2h | 3.00 |
| Realtime | 5.50 |
| Forecast1h No\_Intensity | 5.62 |

### Training Process

Train Loss RMSE: 

Validation Mean Absolute Error:A graph with different colored lines

Description automatically generated

### Conclusion

Significant improvement: Forecast1h reached an MAE of 1.74 (compared to 6.7 for unnormalized data). All normalized networks performed significantly better.  
Training loss over iterations was less choppy.  
Forecast1h performed better than Forecast2h as expected.

## Enhancing the Model with Residual Connections

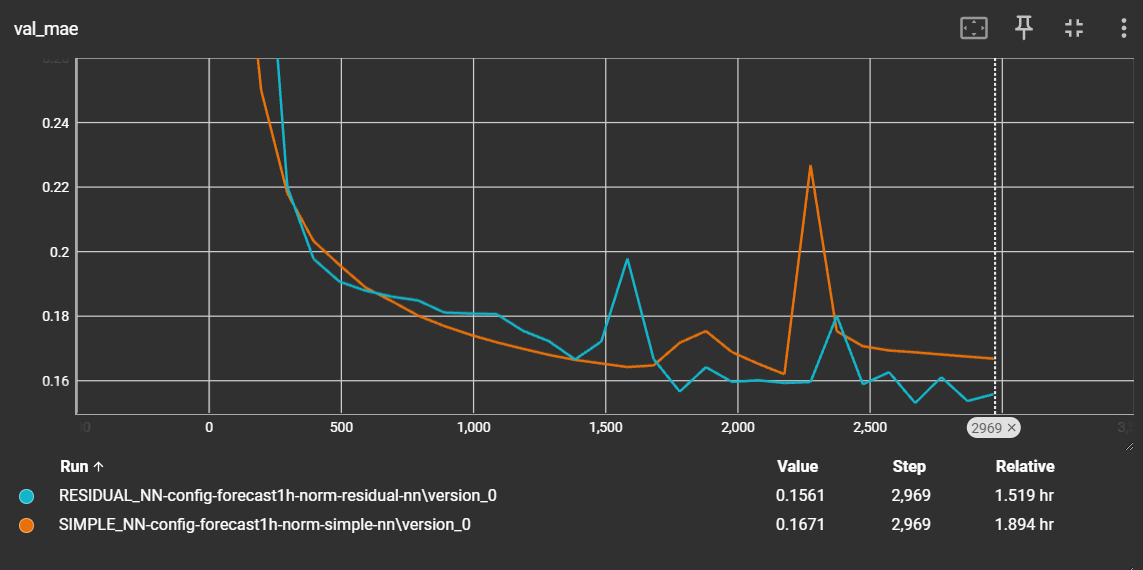
### Experiment Setup

* Training dataset years: 1990-2000
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized Data (as before).
* Added residual connections to the Simple NN from before.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.

### Results

|  |  |
| --- | --- |
| Residual NN Normalized Inputs | Validation Mean Absolute Error |
| Forecast1h | 1.645 |

### Training Process

Validation Mean Absolute Error. Residual NN (blue) vs. SimpleNN (orange):

### Conclusion

Improvement over the Simple NN: Residual NN reached an MAE of 1.645.

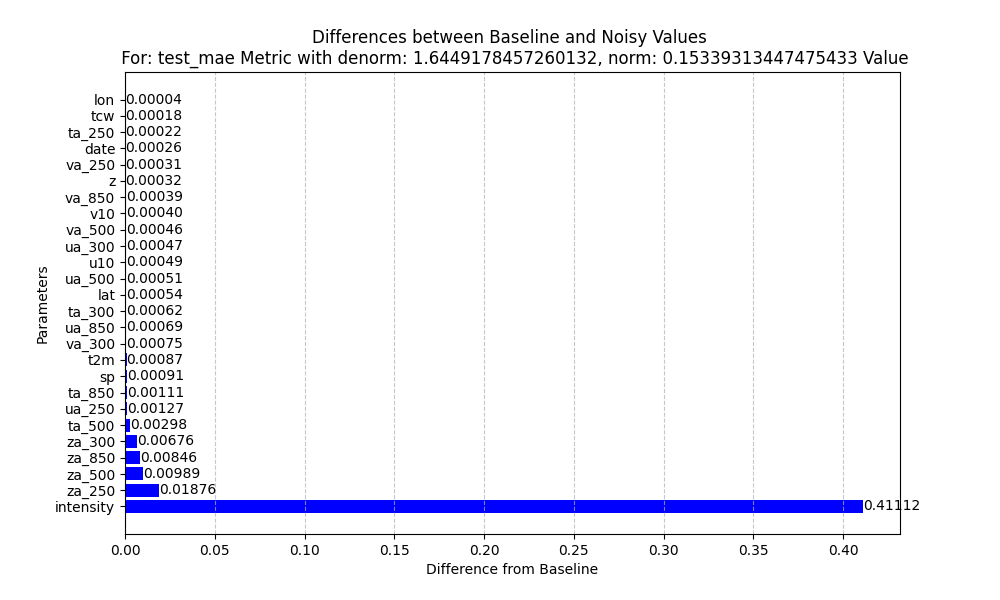
We have now successfully surpassed the Naïve model that only reached a MAE of 1.697 for Forecast1h without lookback.

## Data Processing and Feature Importance

### Experiment Setup

* We wish to determine the importance of each parameter.
* Noise was added to each parameter in the trained networks, and the impact on the MAE for the validation set was measured.

### Results

Forecast1h Residual NN Normalized:

Forecast1h Simple NN Normalized:A graph with numbers and a blue line

Description automatically generated

### Conclusion

* The last intensity value was the most important parameter in networks where it was revealed.
* za was also an important parameter.
* Date and Lon were not beneficial due to their circular nature.

## Circular Unfolding for Circular Parameters

### Experiment Setup

* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* Residual NN (as before).
* Normalized date and lon to the range .
* Added the sin and cos of the normalized date and lon parameters as inputs to the network.
* Added a lookback3h.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.

### Results

|  |  |
| --- | --- |
| Residual NN Normalized Inputs | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 0.98 |

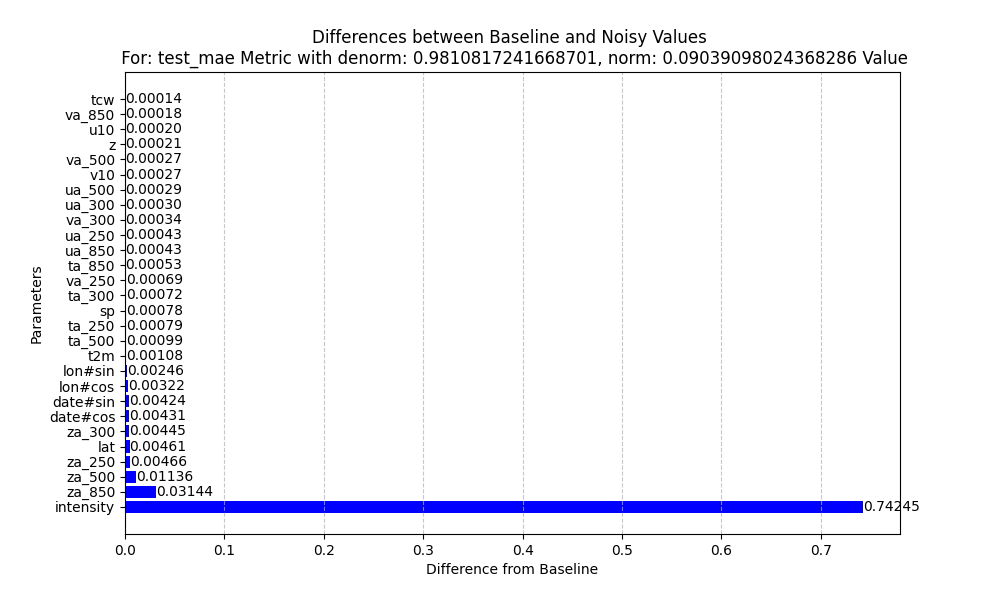
### Training Process

Validation Mean Absolute Error:

A graph with lines and numbers

Description automatically generated

### Feature Importance Analysis



### Conclusion

We have now successfully surpassed the best Naïve model that used a lookback. The best Naïve model used lookback1h only reached a MAE of 1.065 for Forecast1h. We have now been able to reach MAE of 0.98.

Notice how a lot of the parameters seem to be not important still. We will attempt to remove them and see how it affects performance.

## Reducing Unimportant Features

### Experiment Setup

* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* Residual NN (as before).
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Removed all features but: date, lon, lat, za, intensity.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.

### Results

|  |  |
| --- | --- |
| Residual NN Normalized Reduced Inputs | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 0.94 |

### Training Process

Validation Mean Absolute Error:A graph with blue lines

Description automatically generated

### Feature Importance Analysis

A graph with numbers and a blue line

Description automatically generated

### Conclusion

Further improved to 0.94 MAE on the validation set.

We concluded that the other parameters we left out were not important to the network as we were able to beat our best benchmark without them.

## Adding Skip Connection with Naïve Model Output

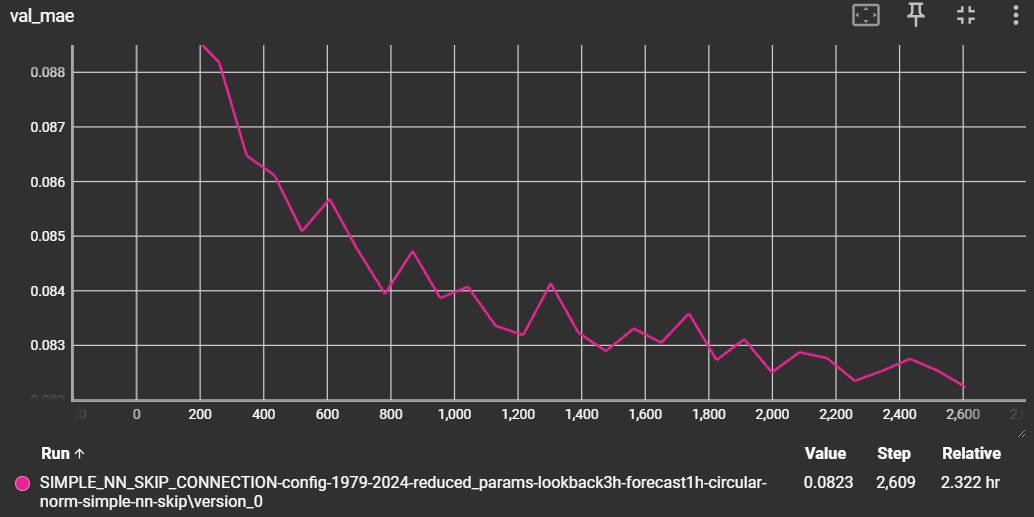
### Experiment Setup

* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* Simple NN.
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Removed all features but: date, lon, lat, za, intensity.
* Added a skip connection to the last linear layer with the Naïve model outputs for lookbacks0-3h.
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.

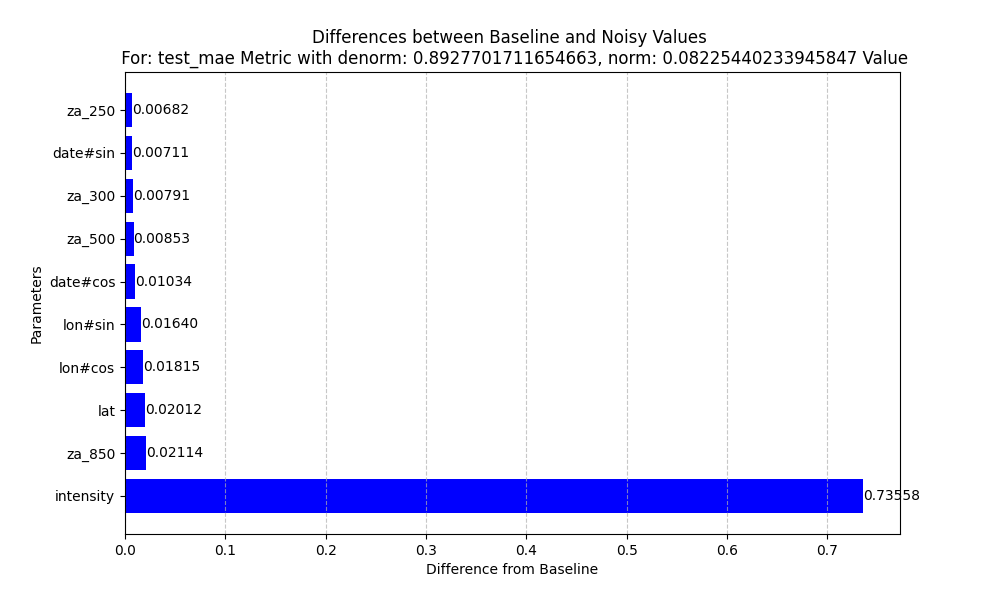
### Results

|  |  |
| --- | --- |
| Simple NN Normalized Reduced Inputs Skip Connection | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 0.89 |

### Training Process

Validation Mean Absolute Error:

### Feature Importance Analysis



### Conclusion

Further improved to 0.89 MAE on the validation set.

**ISSUES WERE DISCOVRED WITH TRAINING DATA.**

**FROM THIS POINT ON, TRAINING DATA WAS FIXED.**

## Adding Skip Connection with Naïve Model Output on Input Data V2

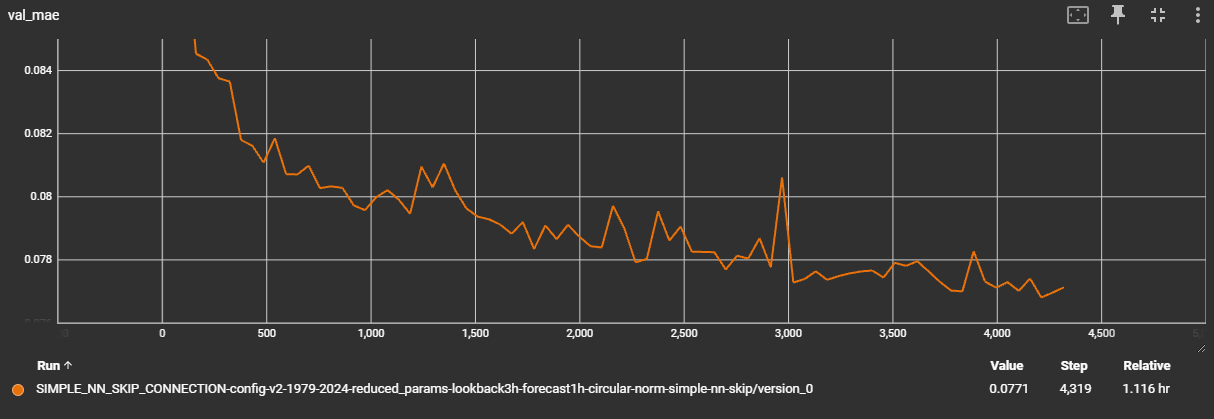
### Experiment Setup

* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* Simple NN.
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Removed all features but: date, lon, lat, za, intensity.
* Added a skip connection to the last linear layer with the Naïve model outputs for lookbacks0-3h.
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.
* Input data V2

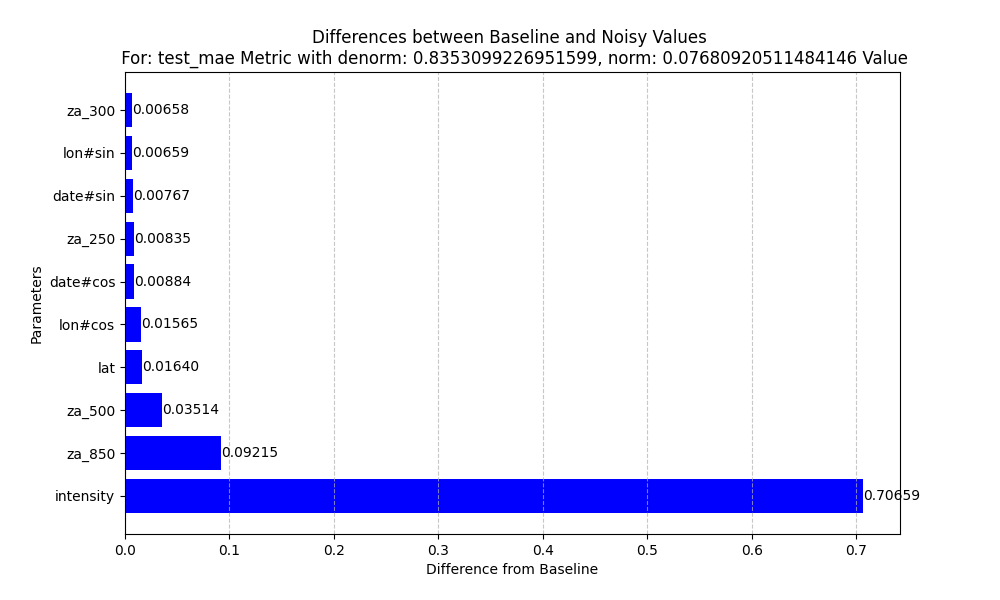
### Results

|  |  |
| --- | --- |
| Simple NN Normalized Reduced Inputs Skip Connection | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 0.83 |

### Training Process

Validation Mean Absolute Error: 

### Feature Importance Analysis



### Conclusion

Further improved to 0.83 MAE on the validation set. This indicates that the parameter fix was effective.

## Skip Connection Naïve Model Output Input Data V2 New Parameters Including M.S.L.

### Experiment Setup

* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* Simple NN.
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Added a skip connection to the last linear layer with the Naïve model outputs for lookbacks0-3h.
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.
* Input data V2
* Added new parameters including M.S.L.

### Results

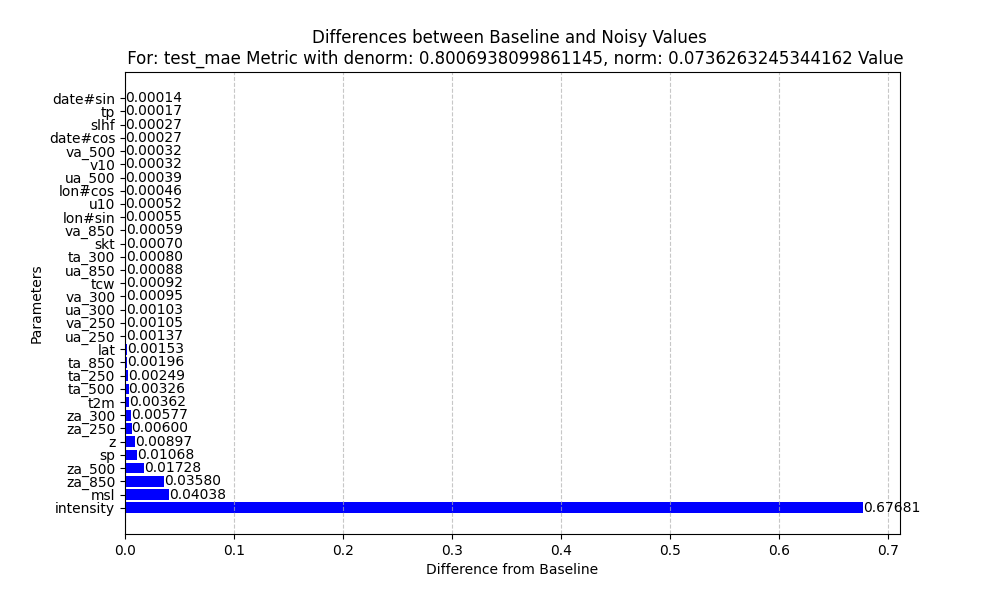
|  |  |
| --- | --- |
| Simple NN Normalized Skip Connection | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 0.80 |

### Training Process

Validation Mean Absolute Error: A graph with blue and orange lines

Description automatically generated

### Feature Importance Analysis



### Conclusion

Further improved to 0.80 MAE on the validation set.

## Skip Connection Naïve Model Output Input Data V2 New Parameters Including M.S.L. 1945-2024

### Experiment Setup

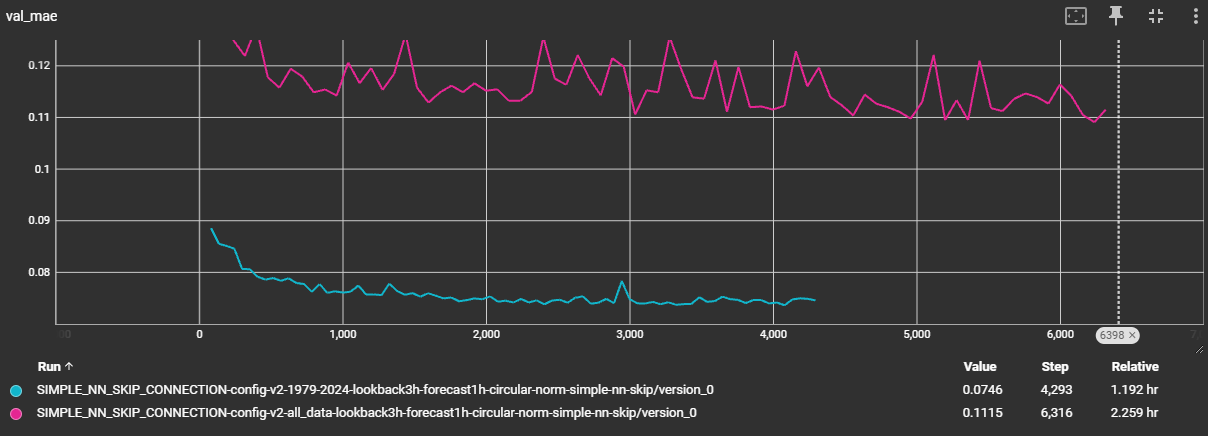
* Training dataset years: 1945-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* Simple NN.
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Added a skip connection to the last linear layer with the Naïve model outputs for lookbacks0-3h.
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0.
* Trained a Residual NN on normalized inputs and tested the Forecast1h version.
* Input data V2
* Added new parameters including M.S.L.

### Results

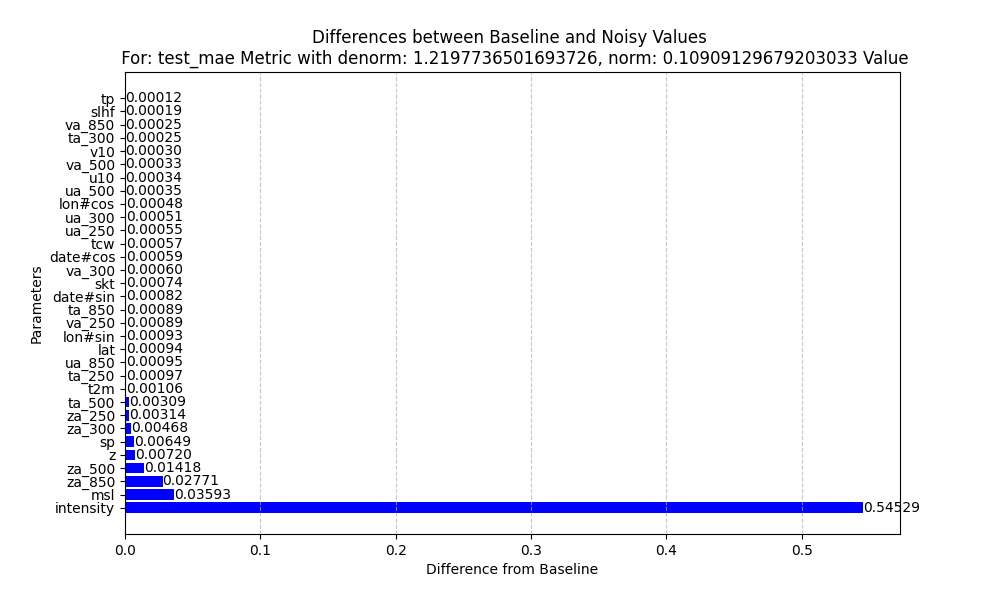
|  |  |
| --- | --- |
| Simple NN Normalized Skip Connection | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 1.21 |

### Training Process

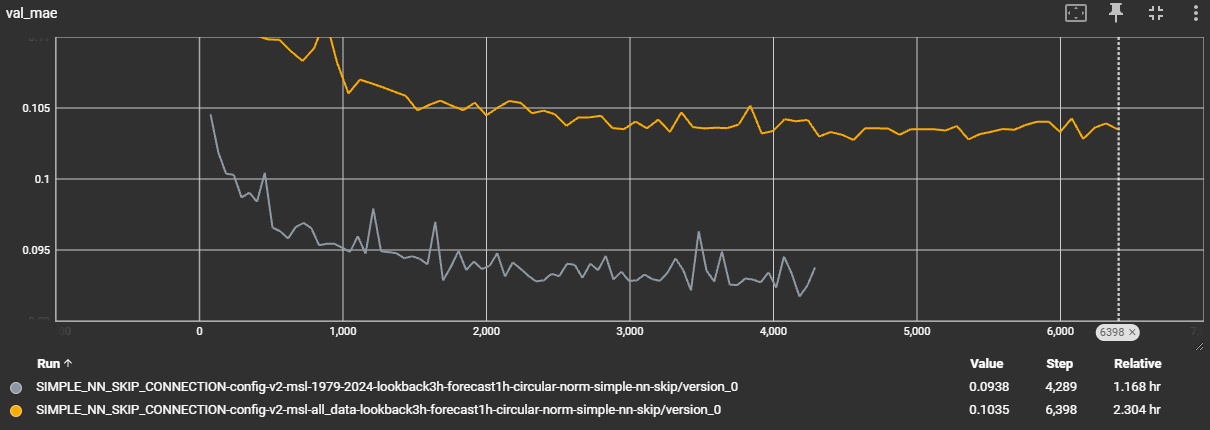
Validation Mean Absolute Error:



### Feature Importance Analysis

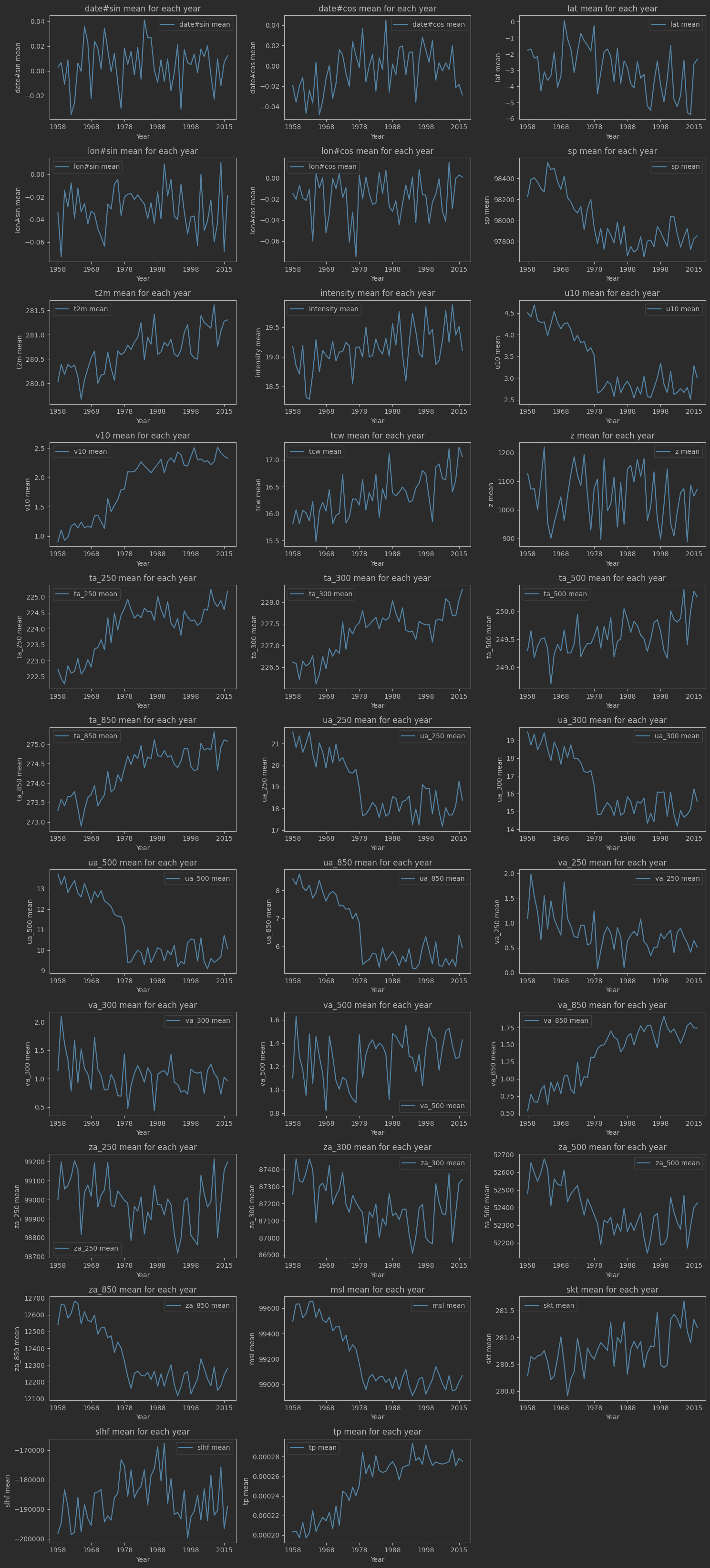


### Conclusion

Got 1.21 MAE on the validation set. This suggests that there are issues with the data before the year 1979. To confirm this suspicion, we ran a similar experiment with the M.S.L. as target instead of the intensity parameter.  
again, we get that the model validation MAE is much worse when the network trains on data between 1945-2024 compared to 1979-2024.

We decided to generate statistics over time for each of the parameters:

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated

## Skip Connection CNN Model Output Input Data V2 New Parameters Including M.S.L.

### Experiment Setup

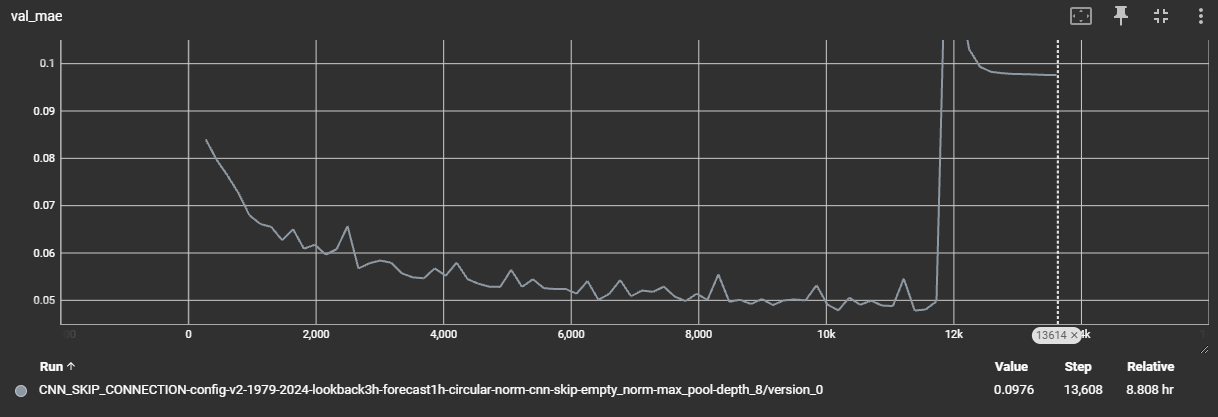
* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* CNN with maxpool layers of depth 8.
* The first 4 layers are cnn->maxpool with stride 2
* The next 4 layers are linear->relu
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Added a skip connection to the last linear layer with the model outputs for lookbacks0-3h. (as before)
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0. (as before)
* Input data V2
* Added new parameters including M.S.L.

### Results

|  |  |
| --- | --- |
| Skip Connection CNN | Validation Mean Absolute Error |
| Lookback3h-Forecast1h | 0.52 |

### Training Process

Validation Mean Absolute Error:



### Feature Importance Analysis

A graph with a blue line

Description automatically generated

### Storm Classification Error

|  |  |  |  |
| --- | --- | --- | --- |
| storm\_classification\_type | test\_mae | denormalized\_test\_mae | num\_storms |
| MEDITERRANEAN | 0.046003 | 0.50029 | 207 |
| SEA | 0.046428 | 0.504907 | 8601 |
| LAND | 0.051779 | 0.563104 | 2651 |
| SOUTH\_HEMISPHERE | 0.045987 | 0.500118 | 5831 |
| NORTH\_HEMISPHERE | 0.049282 | 0.535946 | 5421 |
| NORTH\_ATLANTIC | 0.046952 | 0.510608 | 1375 |
| PACIFIC | 0.047417 | 0.515665 | 2025 |

### Conclusion

Further improved to 0.52 MAE on the validation set over 0.80 MAE on an equivalent center pixel NN.  
This suggests that using the spatial information is in-fact effective.

## MSL Target CNN Skip Connection

### Experiment Setup

* Training dataset years: 1979-2024
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* CNN with maxpool layers of depth 8.
* The first 4 layers are cnn->maxpool with stride 2
* The next 4 layers are linear->relu
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Added a skip connection to the last linear layer with the model outputs for lookbacks0-3h. (as before)
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0. (as before)
* Input data V2
* Added new parameters
* Target is now M.S.L.

### Results

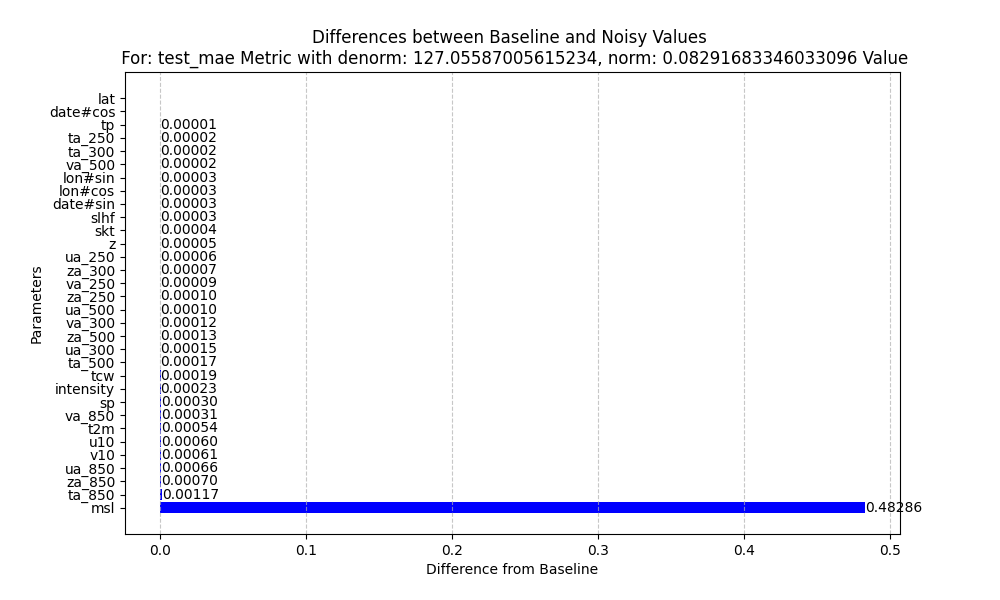
|  |  |
| --- | --- |
|  | Validation Mean Absolute Error |
| CNN Normalized Skip Connection Lookback3h-Forecast1h | 127.0 |
| Naïve Lookback0h-Forecast1h | 242.2 |
| Naïve Lookback1h-Forecast1h | 259.5 |
| Naïve Lookback2h-Forecast1h | 226.1 |
| Naïve Lookback3h-Forecast1h | 232.8 |

### Training Process

Validation Mean Absolute Error: A graph with a line going up

Description automatically generated

### Feature Importance Analysis



### Storm Classification Error

|  |  |  |  |
| --- | --- | --- | --- |
| storm\_classification\_type | test\_mae | denormalized\_test\_mae | num\_storms |
| SEA | 0.082093 | 125.7928 | 8601 |
| SOUTH\_HEMISPHERE | 0.07864 | 120.5025 | 5831 |
| LAND | 0.086904 | 133.1649 | 2651 |
| NORTH\_HEMISPHERE | 0.087634 | 134.2844 | 5421 |
| MEDITERRANEAN | 0.059448 | 91.09399 | 207 |
| NORTH\_ATLANTIC | 0.090234 | 138.2685 | 1375 |
| PACIFIC | 0.086395 | 132.3863 | 2025 |

### Conclusion

MSL CNN performance is about twice as good as any of the naïve models.  
Very low feature importance for any feature that is not MSL.  
Storm classification error is similar on every classification type except MEDITERRANEAN storms where it seems to perform better.

## MSL Target CNN Skip Connection w. Next Storm Step Location Leak 1979-2000

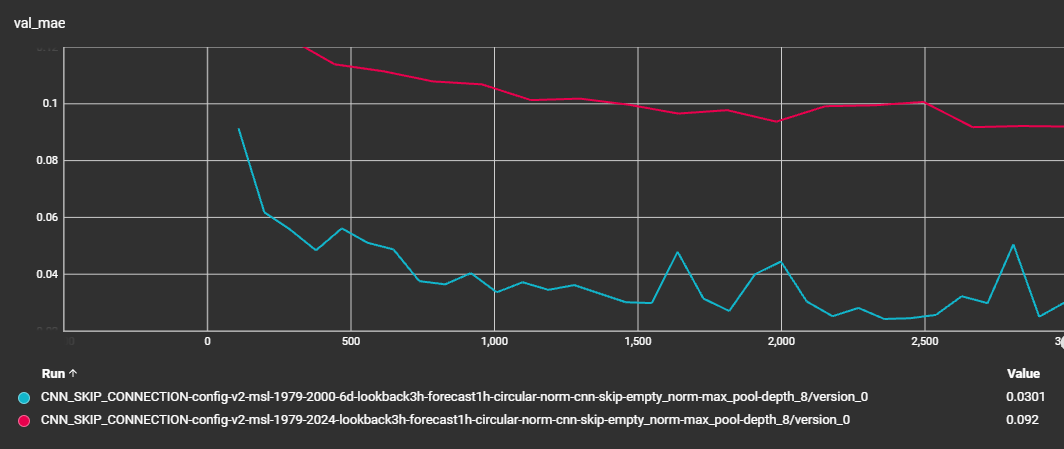
### Experiment Setup

* Training dataset years: 1979-2000
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* CNN with maxpool layers of depth 8.
* The first 4 layers are cnn->maxpool with stride 2
* The next 4 layers are linear->relu
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Added a skip connection to the last linear layer with the model outputs for lookbacks0-3h. (as before)
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0. (as before)
* Input data V2
* Added new parameters
* Target is now M.S.L.
* Added storm location leak- the location of the storm at the next timestep is leaked. Meteorological parameters are also provided for the leaked location as the difference from the previous storm location on these parameters.

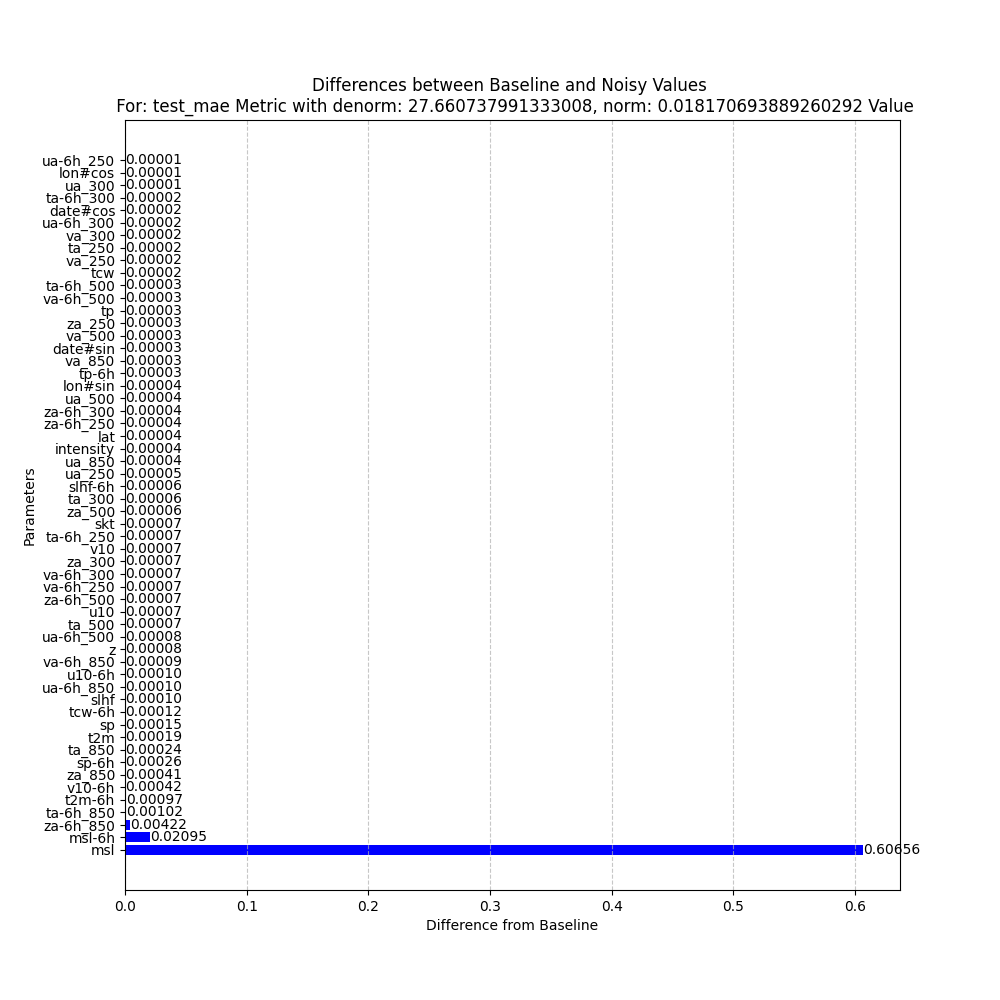
### Results

|  |  |
| --- | --- |
|  | Validation Mean Absolute Error |
| Next Step Leak- CNN Normalized Skip Connection Lookback3h-Forecast1h | 27.66 |
| CNN Normalized Skip Connection Lookback3h-Forecast1h | 127.0 |
| Naïve Lookback0h-Forecast1h | 242.2 |
| Naïve Lookback1h-Forecast1h | 259.5 |
| Naïve Lookback2h-Forecast1h | 226.1 |
| Naïve Lookback3h-Forecast1h | 232.8 |

### Training Process

Validation Mean Absolute Error: 

### Feature Importance Analysis



### Storm Classification Error

|  |  |  |  |
| --- | --- | --- | --- |
| storm\_classification\_type | test\_mae | denormalized\_test\_mae | num\_storms |
| NORTH\_HEMISPHERE | 0.019526 | 29.72395 | 2813 |
| LAND | 0.024825 | 37.79031 | 1359 |
| MEDITERRANEAN | 0.032357 | 49.25582 | 107 |
| PACIFIC | 0.017381 | 26.45819 | 1060 |
| SOUTH\_HEMISPHERE | 0.016171 | 24.61661 | 3038 |
| NORTH\_ATLANTIC | 0.016313 | 24.83269 | 732 |
| SEA | 0.016625 | 25.30768 | 4492 |

### Conclusion

Massive improvement over previous networks. This is made even more promising by the fact that it was only trained on half the data 1979-2000.  
As seen the “training process” chart- after just 1 epoch, the network reaches a loss that is lower than the best loss achieved by the previous best network.  
Parameters from leaked location are placed high in the feature importance chart.  
storms over the MEDITERRANEAN and LAND seem to have poorer performance. This contrasts with the previous model where MEDITERRANEAN storm performance was best. This is possibly due to the low number of samples over the MEDITERRANEAN.

## Intensity Target CNN Skip Connection w. Next Storm Step Location Leak 1979-2000

### Experiment Setup

* Training dataset years: 1979-2000
* Validation was selected randomly. 25% holdout of training. Done deterministically across all experiments.
* Normalized data (as before).
* CNN with maxpool layers of depth 8.
* The first 4 layers are cnn->maxpool with stride 2
* The next 4 layers are linear->relu
* Normalized date and lon to the range (as before.
* Added the sin and cos of the normalized date and lon parameters as inputs to the network (as before).
* Added a lookback3h (as before).
* Added a skip connection to the last linear layer with the model outputs for lookbacks0-3h. (as before)
* Initialized last lin layer weights to be ¼ for the skip connections and 0 for the rest of the neurons. Biases are initialized to 0. (as before)
* Input data V2
* Added new parameters
* Added storm location leak- the location of the storm at the next timestep is leaked. Meteorological parameters are also provided for the leaked location as the difference from the previous storm location on these parameters.

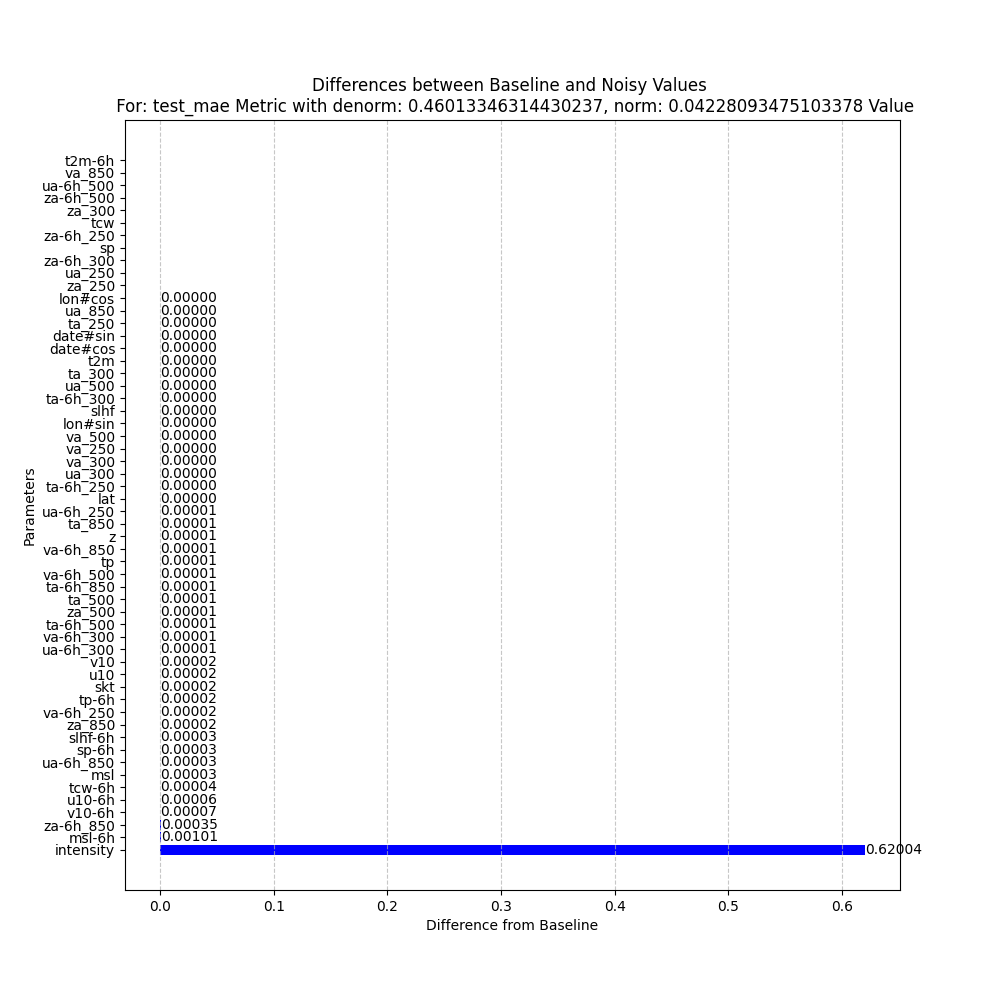
### Results

|  |  |
| --- | --- |
|  | Validation Mean Absolute Error |
| Next Step Leak- CNN Normalized Skip Connection Lookback3h-Forecast1h | 0.46 |
| CNN Normalized Skip Connection Lookback3h-Forecast1h | 0.52 |

### Training Process

Validation Mean Absolute Error: 

### Feature Importance Analysis



### Storm Classification Error

|  |  |  |  |
| --- | --- | --- | --- |
| storm\_classification\_type | test\_mae | denormalized\_test\_mae | num\_storms |
| NORTH\_HEMISPHERE | 0.043107 | 0.469118 | 2813 |
| NORTH\_ATLANTIC | 0.040831 | 0.444358 | 732 |
| SEA | 0.041403 | 0.450583 | 4492 |
| PACIFIC | 0.042786 | 0.465633 | 1060 |
| MEDITERRANEAN | 0.036223 | 0.394206 | 107 |
| SOUTH\_HEMISPHERE | 0.042006 | 0.457145 | 3038 |
| LAND | 0.045229 | 0.492216 | 1359 |

### Conclusion

Slight improvement over previous networks when only trained on half the data 1979-2000.