

# Random Forest Simulated PID Test on Real Data Transfer Learning (FEDA) Comparison Jan2021 Good Data Only

February 11, 2021

```
[1]: from sklearn.ensemble import RandomForestRegressor
from math import nan
import math
import numpy as np
import pandas as pd
import csv
import pdb
import os
import matplotlib.pyplot as plt
import timeit
import feda # feda.py must be in same directory
```

```
[2]: def load_data(directory, target_var):
    full_data = []
    if directory[-1] != '/':
        directory = directory + '/'
    for root,dir,files in os.walk(directory):
        for file in files:
            if file.endswith(".csv"):
                filepath = directory + file
                print('Loading: ', file)
                csv_data = np.genfromtxt(filepath, delimiter=',')
                csv_data = np.delete(csv_data, 0, 0)
                full_data.append(csv_data)
    try:
        features = np.vstack(full_data)
        ground_truth = features[:,target_var]
        features = np.delete(features, target_var, axis=1)
    except:
        print("Loading Error: Data not found")
    return features, ground_truth, full_data
```

```
[3]: def load_data_cv(directory, target_var):
    print("Loading data")
```

```

full_data = []
file_names = []
ground_truth = []
features = []
if directory[-1] != '/':
    directory = directory + '/'
for root,dir,files in os.walk(directory):
    for file in files:
        if file.endswith(".csv"):
            filepath = directory + file
            print('Loading: ', file)
            csv_data = np.genfromtxt(filepath, delimiter=',')
            csv_data = np.delete(csv_data, 0, 0)
            full_data.append(csv_data)
            file_names.append(file)
for file in range(0,len(full_data)):
    ground_truth.append(full_data[file][:,target_var])
    features.append(np.delete(full_data[file], target_var, axis=1))

return features, ground_truth, full_data, file_names

```

```

[4]: truth_index = 1
start = timeit.default_timer()
source_features, _, source_data = load_data('/d/git/heat_chamber_code/
↳Updated_Dataset_Jan_2021/Formatted_Data/Sim_Data/', truth_index)
trans_source_features, source_truth=feda.source_transform(source_data,
↳truth_index)

```

```

Loading: NewNewAluminum_Thick_0_-1_0_-1_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_-2_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_-3_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_-4_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_-5_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_-6_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_-7_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_0_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_1_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_2_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_3_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_4_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_5_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_0_6_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_1_-1_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_1_-2_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_1_-3_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_1_-4_formatted.csv
Loading: NewNewAluminum_Thick_0_-1_1_-5_formatted.csv

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

Loading: NewNewSteel_Thin_0_1_3_6_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-1_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-2_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-3_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-4_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-5_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-6_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_-7_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_0_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_1_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_2_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_3_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_4_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_5_formatted.csv
Loading: NewNewSteel_Thin_0_1_4_6_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-1_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-2_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-3_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-4_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-5_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-6_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_-7_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_0_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_1_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_2_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_3_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_4_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_5_formatted.csv
Loading: NewNewSteel_Thin_0_1_5_6_formatted.csv

```

```

[5]: truth_index = 1
start = timeit.default_timer()
cv_features, cv_truth, cv_full_data, cv_file_names = load_data_cv('/d/git/
↳heat_chamber_code/Updated_Dataset_Jan_2021/Formatted_Data/Experimental_Data/
↳Good_Single_Temp/', truth_index)

```

```

Loading data
Loading: NewNewRun2_Aluminum_Thick.csv
Loading: NewNewRun2_Steel_Thin.csv
Loading: NewNewRun3_Steel_Thick.csv
Loading: NewNewRun_1_Aluminum_Thin_mod.csv

```

# 1 Transfer Learning Using Frustratingly Easy Domain Adaptation (FEDA)

```
[6]: trans_cv_features = []
trans_cv_truth = []
for i in range(len(cv_full_data)):
    target_features, target_truth=feda.target_transform(cv_full_data[i],
    ↪truth_index)
    trans_cv_features.append(target_features)
    trans_cv_truth.append(target_truth)
```

```
[7]: train_set_features = []
train_set_truth = []
train_features = []
train_truth = []
num_runs = len(trans_cv_features)
# Build leave one hold out training sets
for test_run in range(0,num_runs):
    for train_run in range(0, num_runs):
        if test_run != train_run:
            train_features.append(trans_cv_features[train_run])
            train_truth.append(trans_cv_truth[train_run])

    train_features.append(trans_source_features)
    train_truth.append(source_truth)
    train_set_truth.append(np.hstack(train_truth))
    train_set_features.append(np.vstack(train_features))
    train_truth = []
    train_features = []
```

```
[8]: forest = RandomForestRegressor(random_state=42)
trans_score= []
trans_prediction_array = []
trans_truth_array = []
trans_error_array = []
for run in range(0,num_runs):
    forest.fit(train_set_features[run], train_set_truth[run])
    trans_score.append(forest.score(trans_cv_features[run],
    ↪trans_cv_truth[run]))
    run_predictions = []
    run_truth = []
    temp_err = []
    index = 0
    for sample in range(0, len(trans_cv_features[run])):
        prediction = forest.predict([trans_cv_features[run][sample]])
        run_predictions.append(prediction)
```



```

        temp_err.append(100*abs((trans_cv_truth[run][sample] - prediction)/
↪trans_cv_truth[run][sample]))
        index = index + 1
    run_predictions = np.vstack(run_predictions)
    trans_error_array.append(np.vstack(temp_err))
    trans_prediction_array.append(run_predictions)
    trans_truth_array.append(trans_cv_truth[run])

```

```

/home/ubuntu/anaconda3/lib/python3.7/site-
packages/sklearn/ensemble/forest.py:245: FutureWarning: The default value of
n_estimators will change from 10 in version 0.20 to 100 in 0.22.
  "10 in version 0.20 to 100 in 0.22.", FutureWarning)

```

```

[9]: average_score = str(np.average(trans_score))
print('Transfer Learning Cross validation score: ' + average_score)

```

Transfer Learning Cross validation score: 0.8436466150539836

## 2 Augmenting Real Data with Simulated Data without using FEDA

```

[10]: train_set_features = []
train_set_truth = []
train_features = []
train_truth = []
num_runs = len(cv_features)
# Build leave one hold out training sets
for test_run in range(0,num_runs):
    for train_run in range(0, num_runs):
        if test_run != train_run:
            train_features.append(cv_features[train_run])
            train_truth.append(cv_truth[train_run])

    train_features.append(source_features)
    train_truth.append(source_truth)
    train_set_truth.append(np.hstack(train_truth))
    train_set_features.append(np.vstack(train_features))
    train_truth = []
    train_features = []

```

```

[11]: forest = RandomForestRegressor(random_state=42)
score = []
prediction_array = []
truth_array = []

```

```

error_array = []
for run in range(0,num_runs):
    forest.fit(train_set_features[run], train_set_truth[run])
    score.append(forest.score(cv_features[run], cv_truth[run]))
    run_predictions = []
    run_truth = []
    temp_err = []
    index = 0
    for sample in range(0, len(cv_features[run])):
        prediction = forest.predict([cv_features[run][sample]])
        run_predictions.append(prediction)
        temp_err.append(100*abs((cv_truth[run][sample] - prediction)/
↪cv_truth[run][sample]))
        index = index + 1
    run_predictions = np.vstack(run_predictions)
    error_array.append(np.vstack(temp_err))
    prediction_array.append(run_predictions)
    truth_array.append(cv_truth[run])

```

/home/ubuntu/anaconda3/lib/python3.7/site-packages/sklearn/ensemble/forest.py:245: FutureWarning: The default value of n\_estimators will change from 10 in version 0.20 to 100 in 0.22.  
 "10 in version 0.20 to 100 in 0.22.", FutureWarning)

```

[12]: average_score = str(np.average(score))
print('Cross validation score: ' + average_score)

```

Cross validation score: 0.6256452064484292

```

[13]: for i in range(0,num_runs):
    trans_score_str = str(trans_score[i])
    score_str = str(score[i])
    print('Transfer Learning Score: ' + trans_score_str)
    print('Standard Scores: ' + score_str)
    plt.figure
    plt.plot(range(len(truth_array[i])), trans_truth_array[i], label="Ground_
↪truth")
    plt.plot(range(len(truth_array[i])), trans_cv_features[i][:,4],
↪label="Outer temp")
    plt.plot(range(len(truth_array[i])), trans_prediction_array[i], label="TL_
↪Prediction", alpha=0.5)
    plt.plot(range(len(truth_array[i])), prediction_array[i], 'r',
↪label="Prediction", alpha=0.3)
    plt.legend(loc="lower right")
    plt.xlabel('Time (s)')
    plt.ylabel('Temperature ($\degree$C)')

```

```

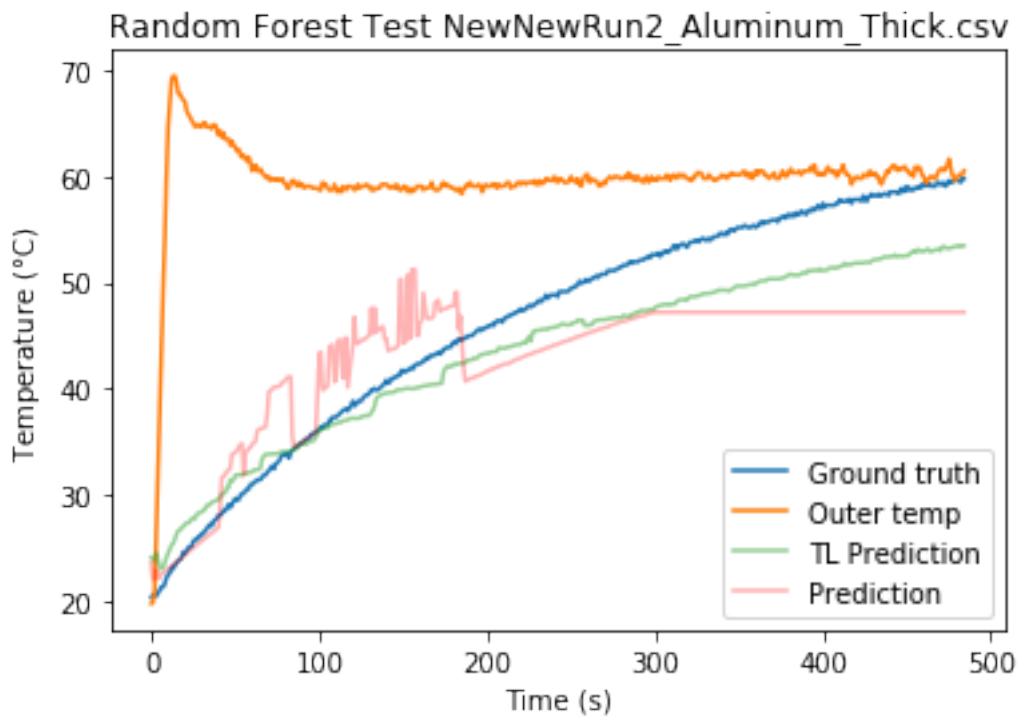
plt.title('Random Forest Test ' + cv_file_names[i])
plt.show()
plt.figure
plt.plot(range(len(trans_error_array[i])), trans_error_array[i],
label="Transfer Learning")
plt.plot(range(len(error_array[i])), error_array[i], label="Standard
Model", alpha=0.5, )
plt.legend(loc="lower right")
plt.xlabel('Time (s)')
plt.ylabel('% Error')
plt.title('Random Forest %Error Test ' + cv_file_names[i])
plt.show()

stop = timeit.default_timer()
print('Time: ', stop - start)

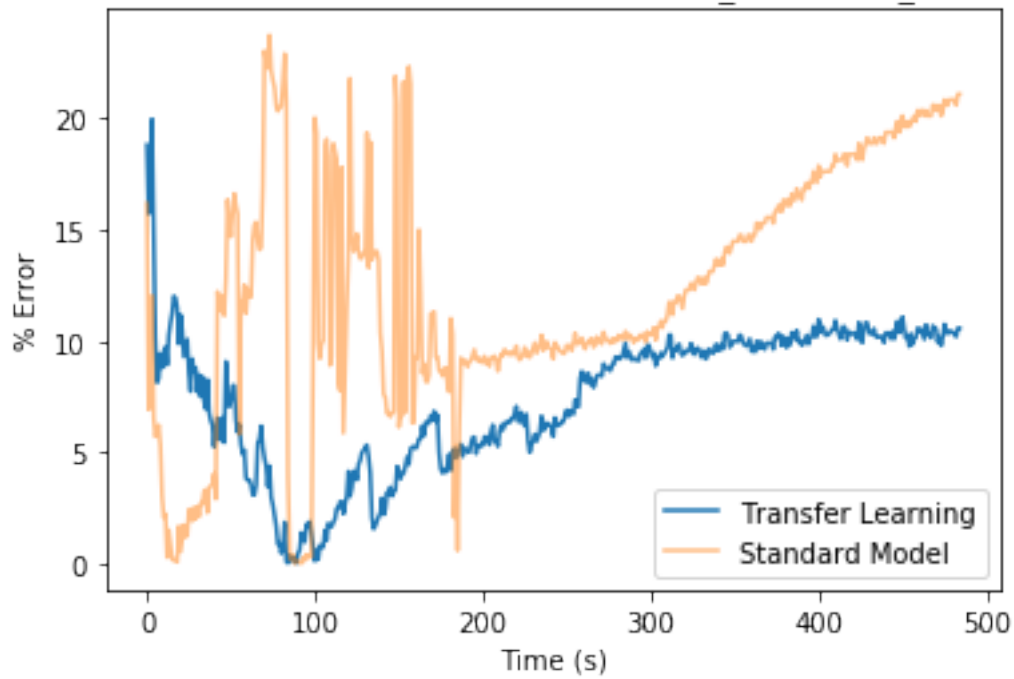
```

Transfer Learning Score: 0.8571577510109802

Standard Scores: 0.593607656284419



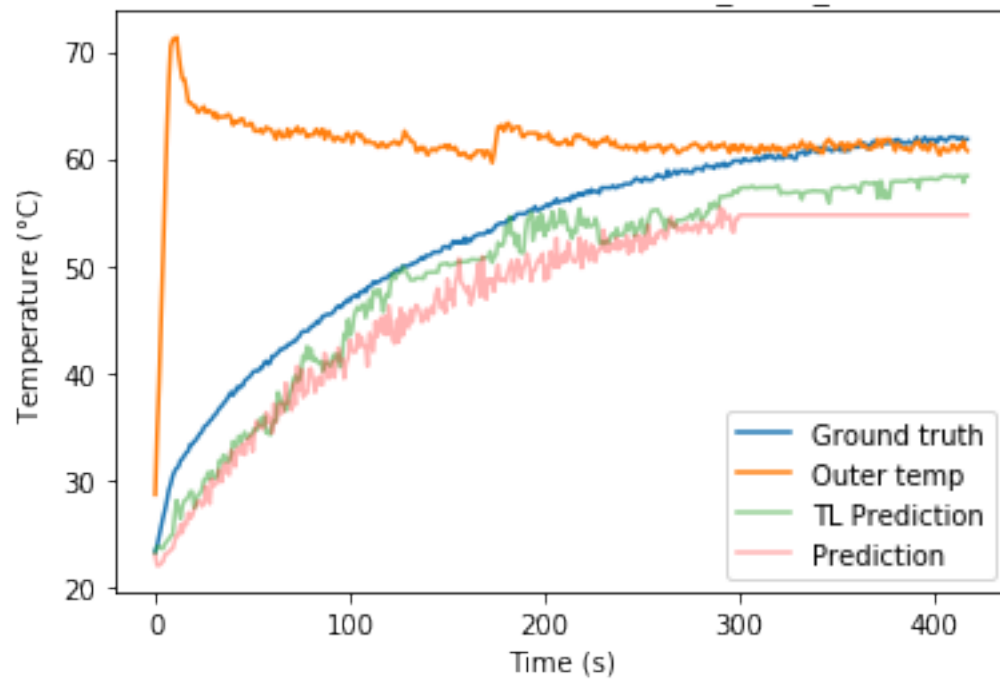
Random Forest %Error Test NewNewRun2\_Aluminum\_Thick.csv

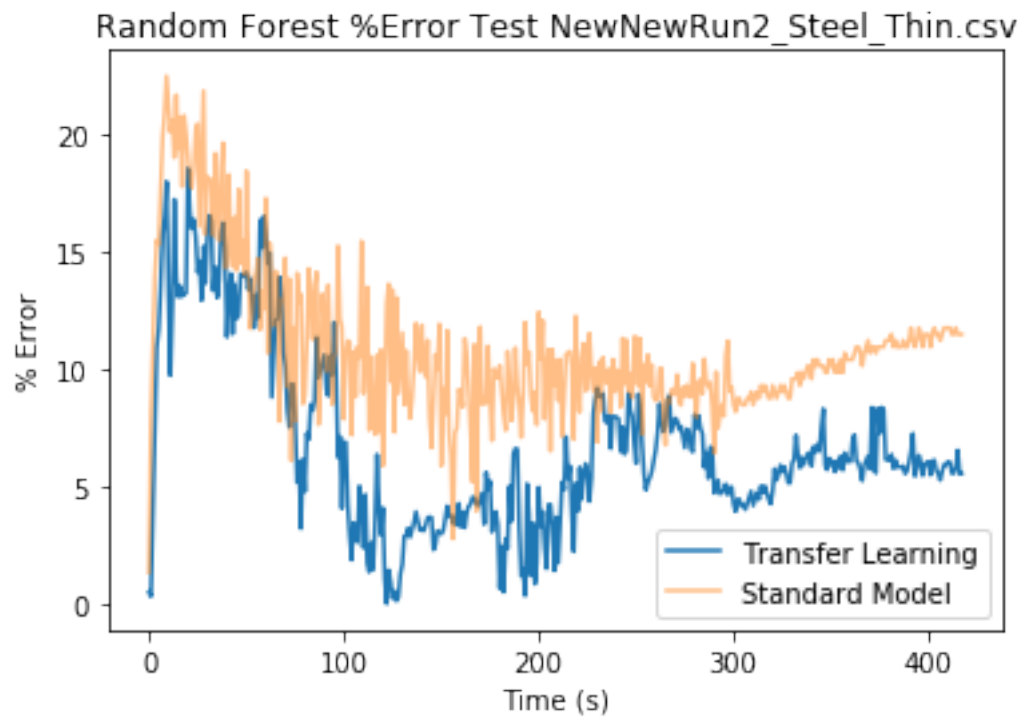


Transfer Learning Score: 0.8423782027582618

Standard Scores: 0.6151965333829744

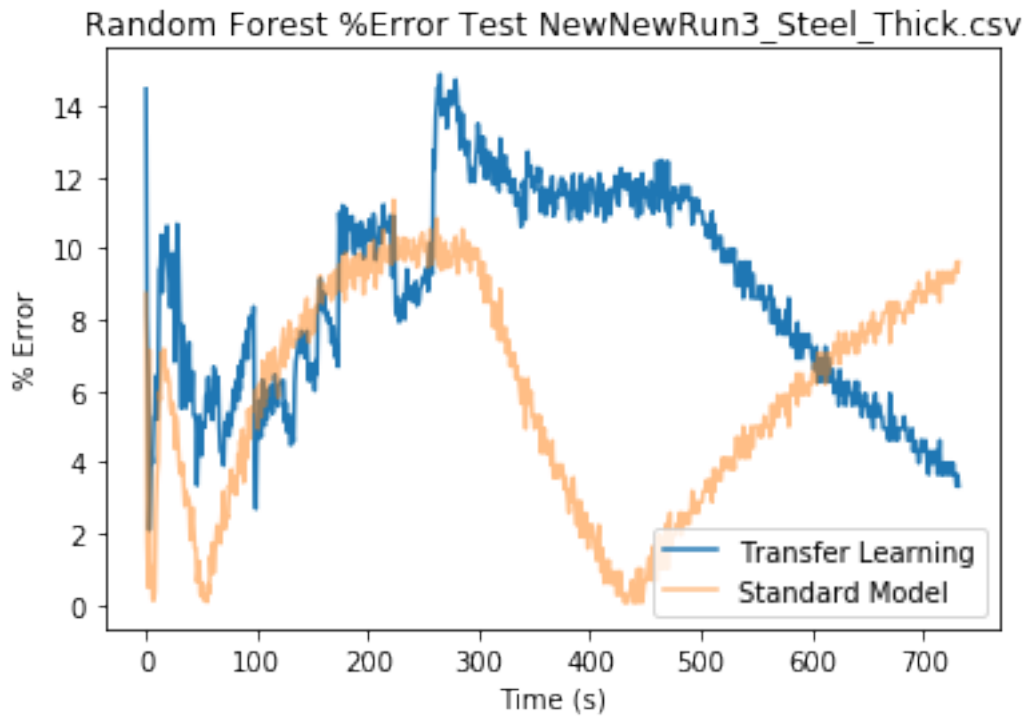
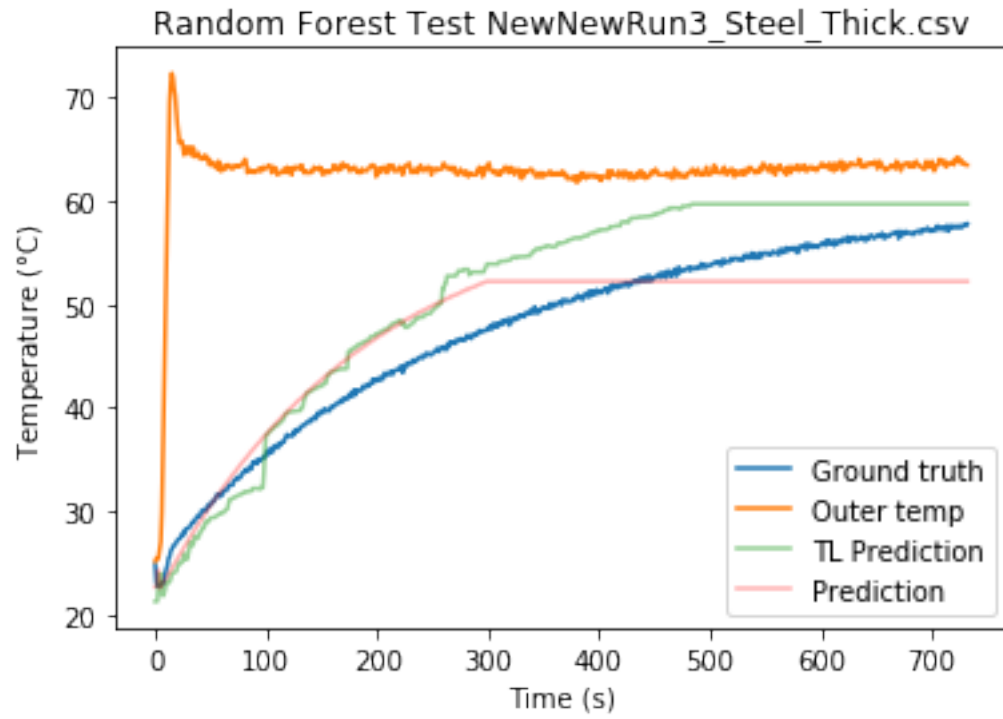
Random Forest Test NewNewRun2\_Steel\_Thin.csv



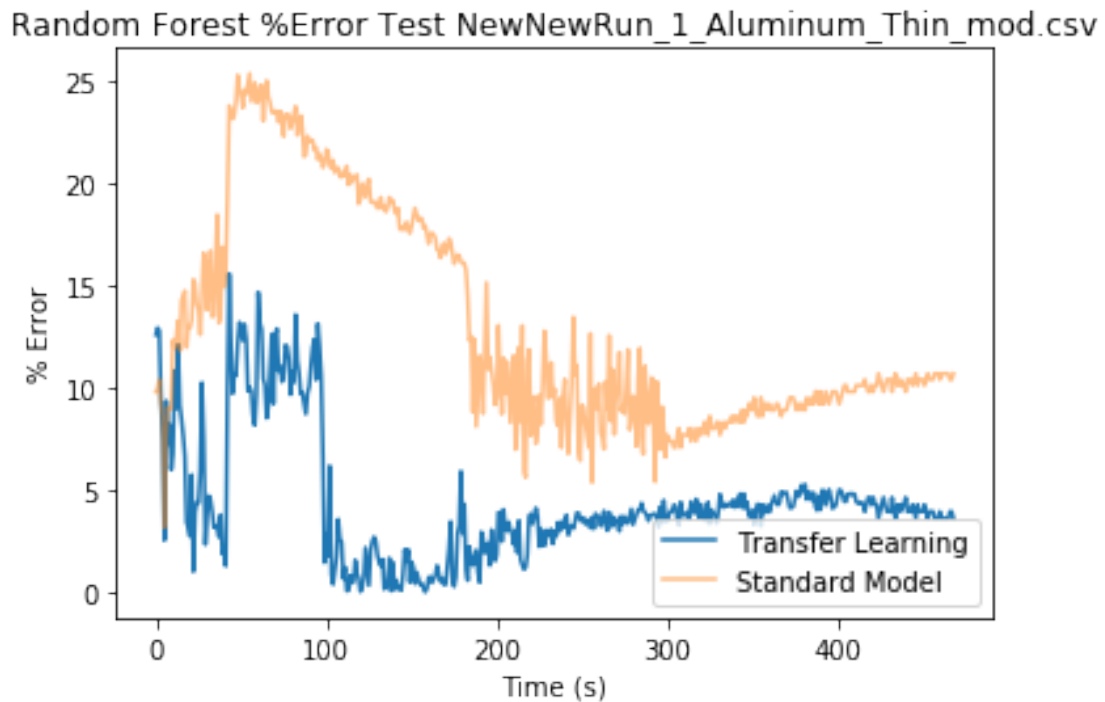
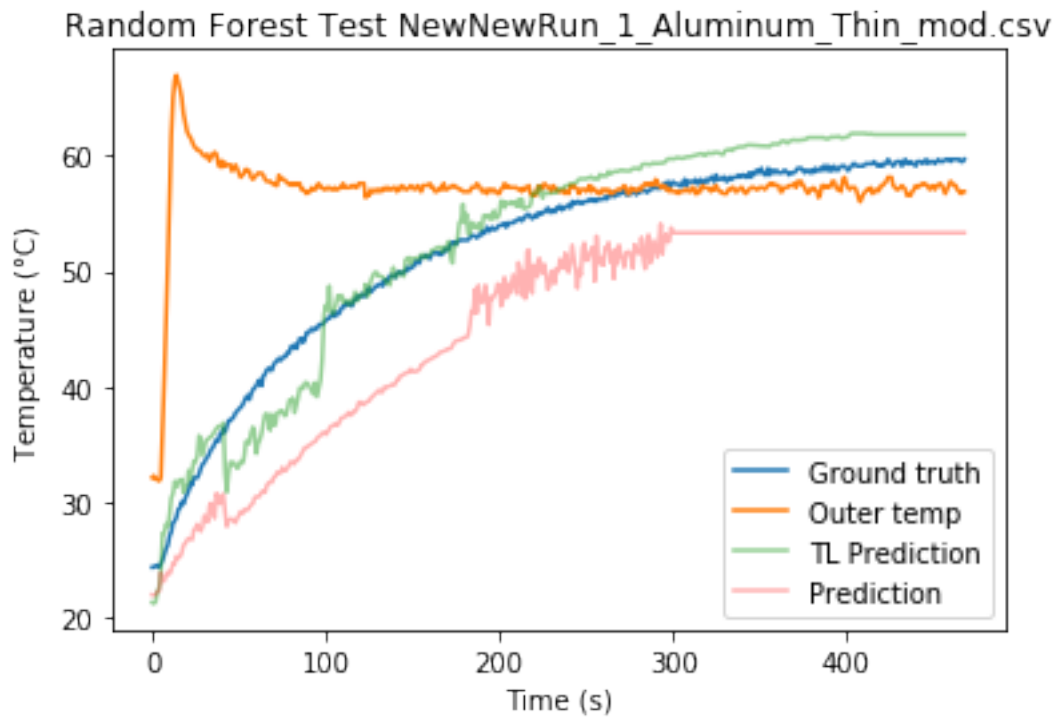


Transfer Learning Score: 0.7511023592880528

Standard Scores: 0.8716490254339547



Transfer Learning Score: 0.9239481471586398  
Standard Scores: 0.4221276106923687



Time: 679.239697