

Final Project

By Andy/Andres

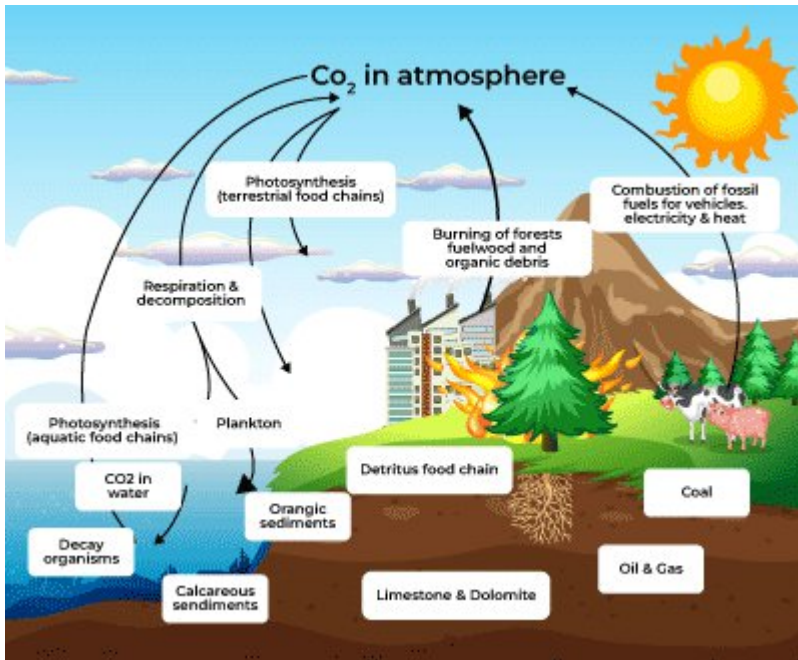
Domain Knowledge

Biological Carbon Cycle

Air_sea Co₂ Exchange

Mixed Layer Depth

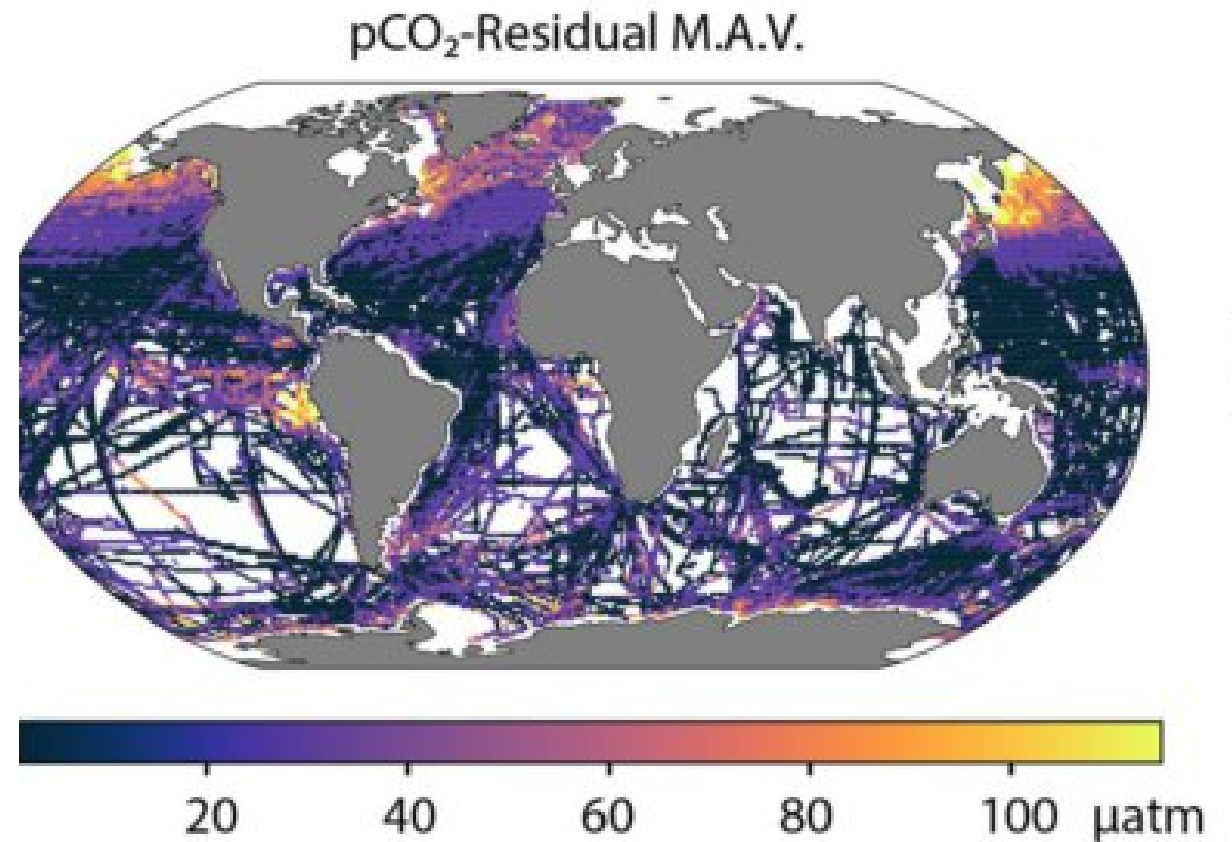
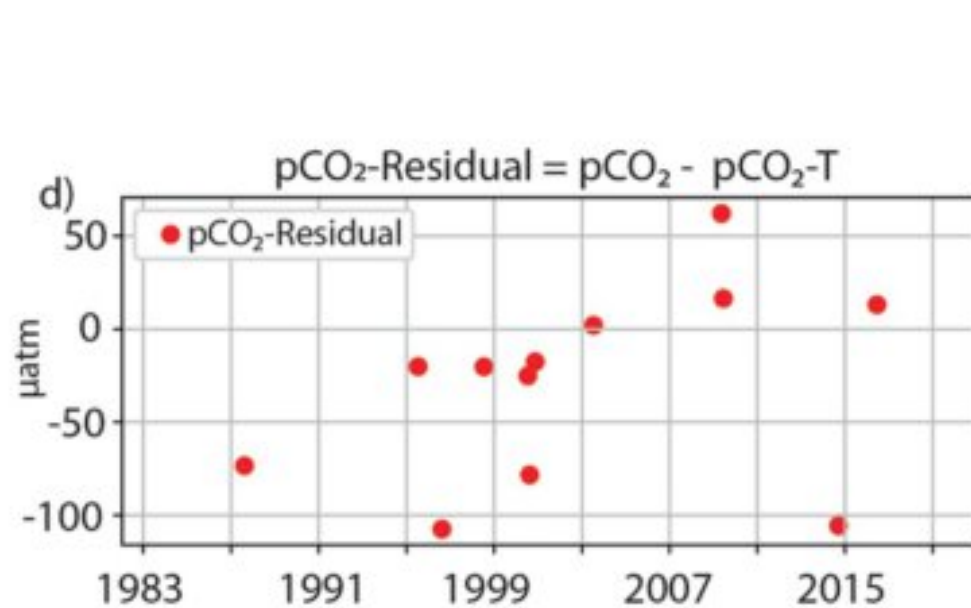
Anthropogenic Co₂ Influence



Reference:

<https://www.geeksforgeeks.org/carbon-cycle-steps/>

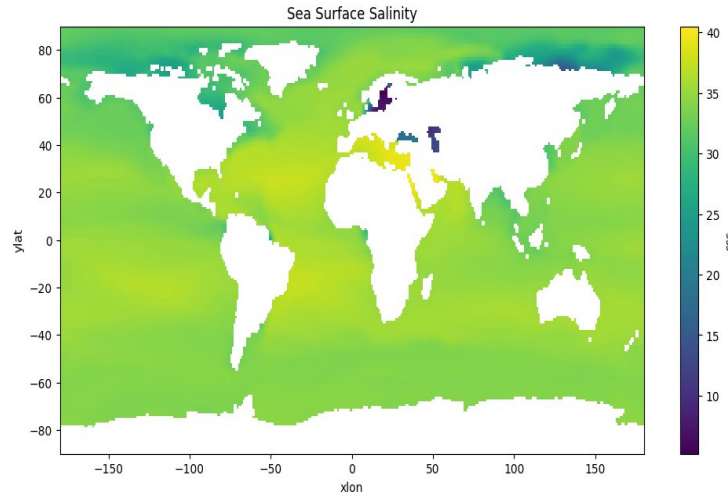
Residual Concept from Research Paper



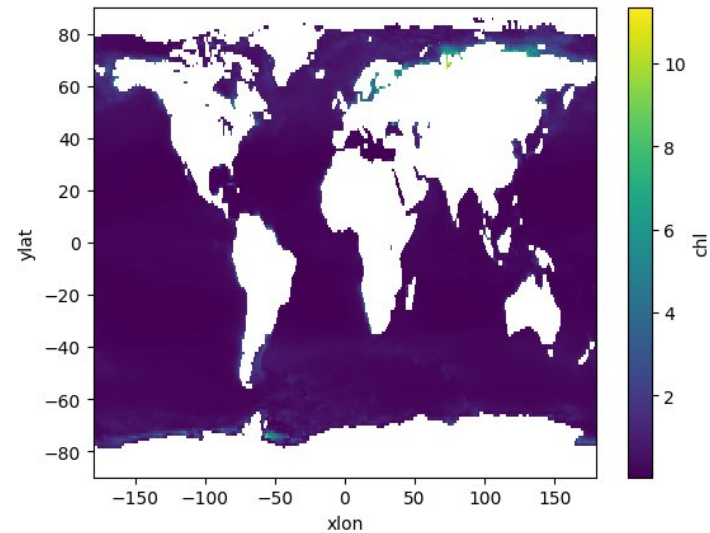
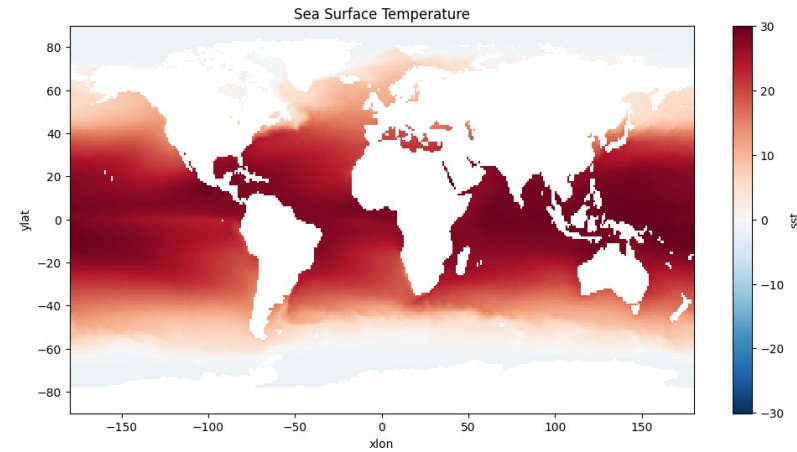
Ref: J Adv Model Earth Syst - 2022 - Bennington - Explicit Physical Knowledge in Machine Learning for Ocean Carbon Flux

Biogeochemical Features Used in Training the Model

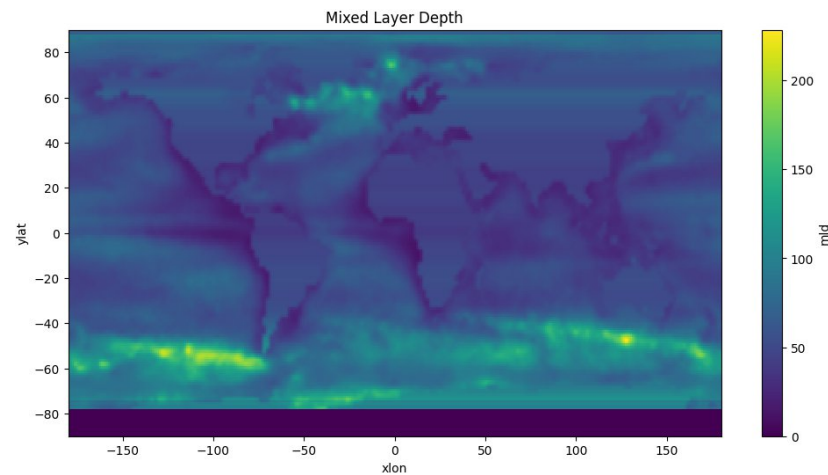
Sea Surface Salinity



Sea Surface Temperature



chlorophyll concentration



Mixed Layer Depth

Results Achieved

```
Fitting 5 folds for each of 50 candidates, totalling 250 fits
```

```
Best Parameters: {'xgb__subsample': 0.8, 'xgb__reg_lambda': 20, 'xgb__reg_alpha':  
20, 'xgb__n_estimators': 700, 'xgb__min_child_weight': 5, 'xgb__max_depth': 9,  
'xgb__learning_rate': 0.2, 'xgb__colsample_bytree': 0.8}
```

```
Best R2 Score: 0.8802070097423262
```

```
Cross-Validation Results:
```

```
Train R2 Scores: [0.98878014 0.98882945 0.98871028 0.98878793 0.98901093]
```

```
Test R2 Scores: [0.88575497 0.88148519 0.88404932 0.87750481 0.87224075]
```

```
Mean Train R2: 0.9888237461398608
```

```
Mean Test R2: 0.8802070097423262
```

Feature engineering and results achieved

```
```python
features.reset_index(inplace=True)

features['sst_lat_trend'] = features['sst'] * features['B']

features['sss_north_south'] = features['sss'] * features['A']

```
```

```
Mean Test Score: 0.8812414022690774 Mean Train Score: 0.9895811420737761
```

Performance Scores by Region and Season

MSE by Region:

| region | |
|--------|------------|
| Middle | 121.755104 |
| North | 300.900456 |
| South | 190.350860 |

R2 Score by Region:

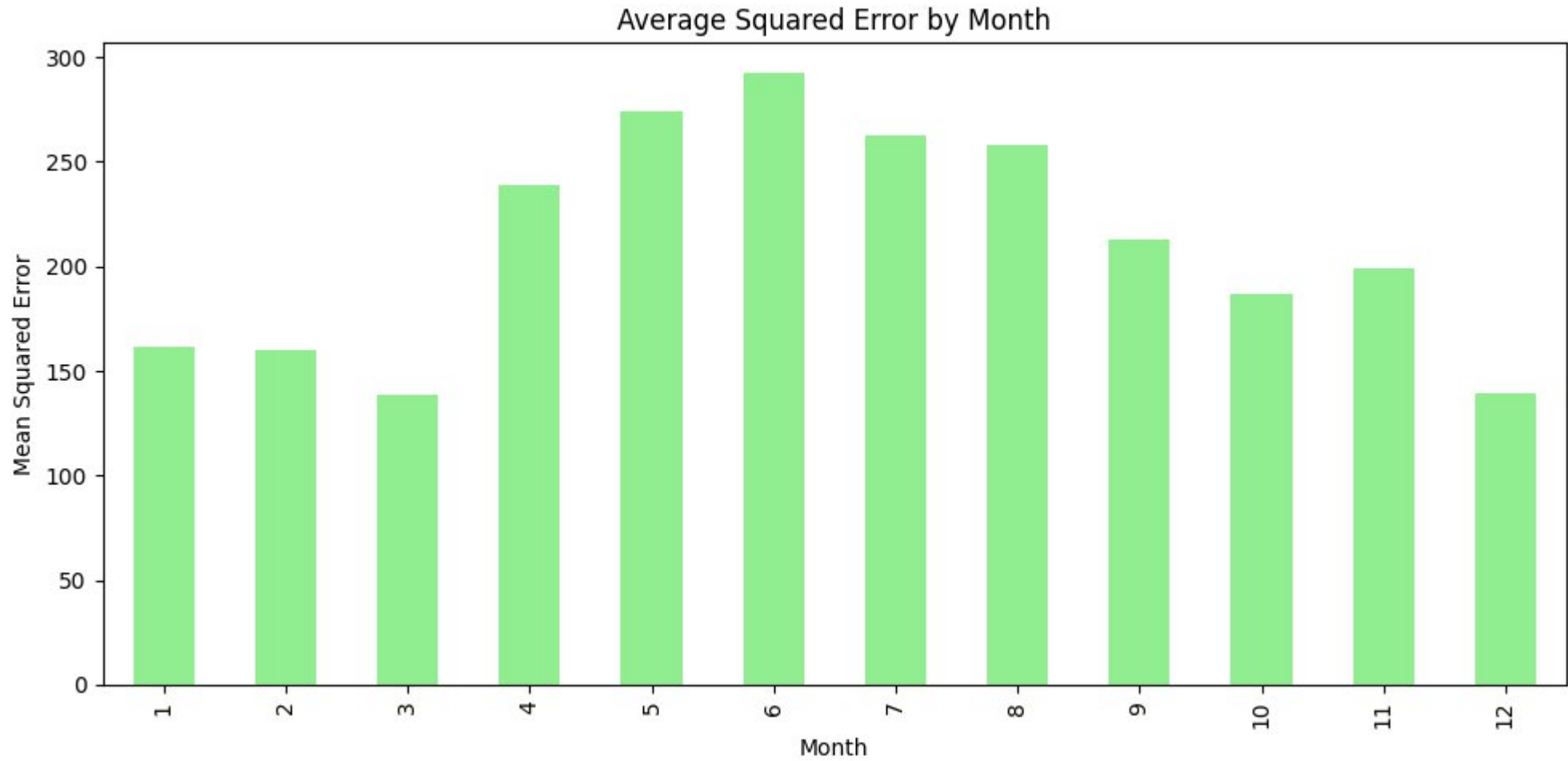
| | |
|----------------|-------|
| R2 for South: | 0.790 |
| R2 for Middle: | 0.875 |
| R2 for North: | 0.853 |

MSE by Season:

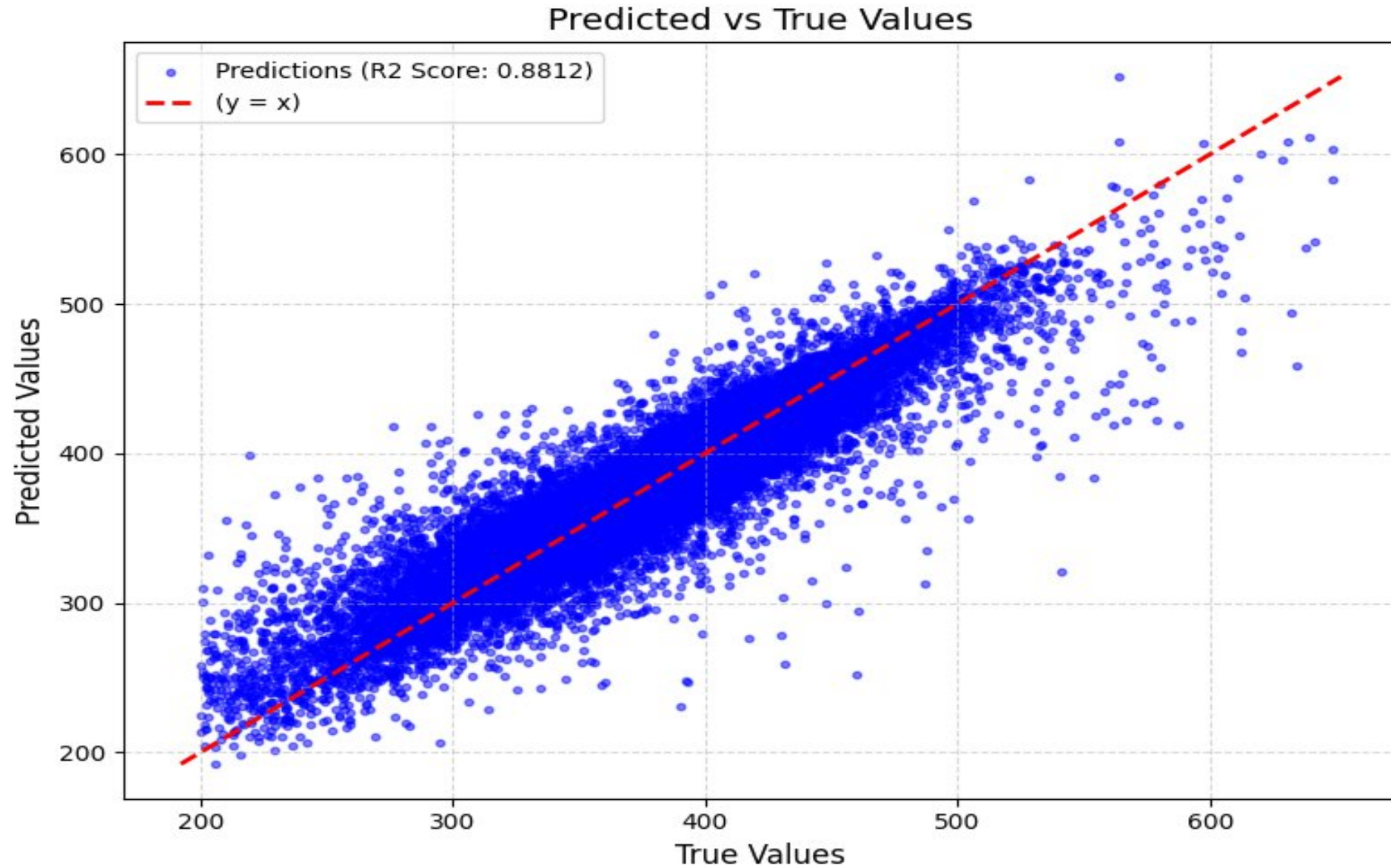
| season | |
|--------|------------|
| Autumn | 192.623131 |
| Spring | 226.326814 |
| Summer | 279.488129 |
| Winter | 99.990688 |

R2 Score by Season:

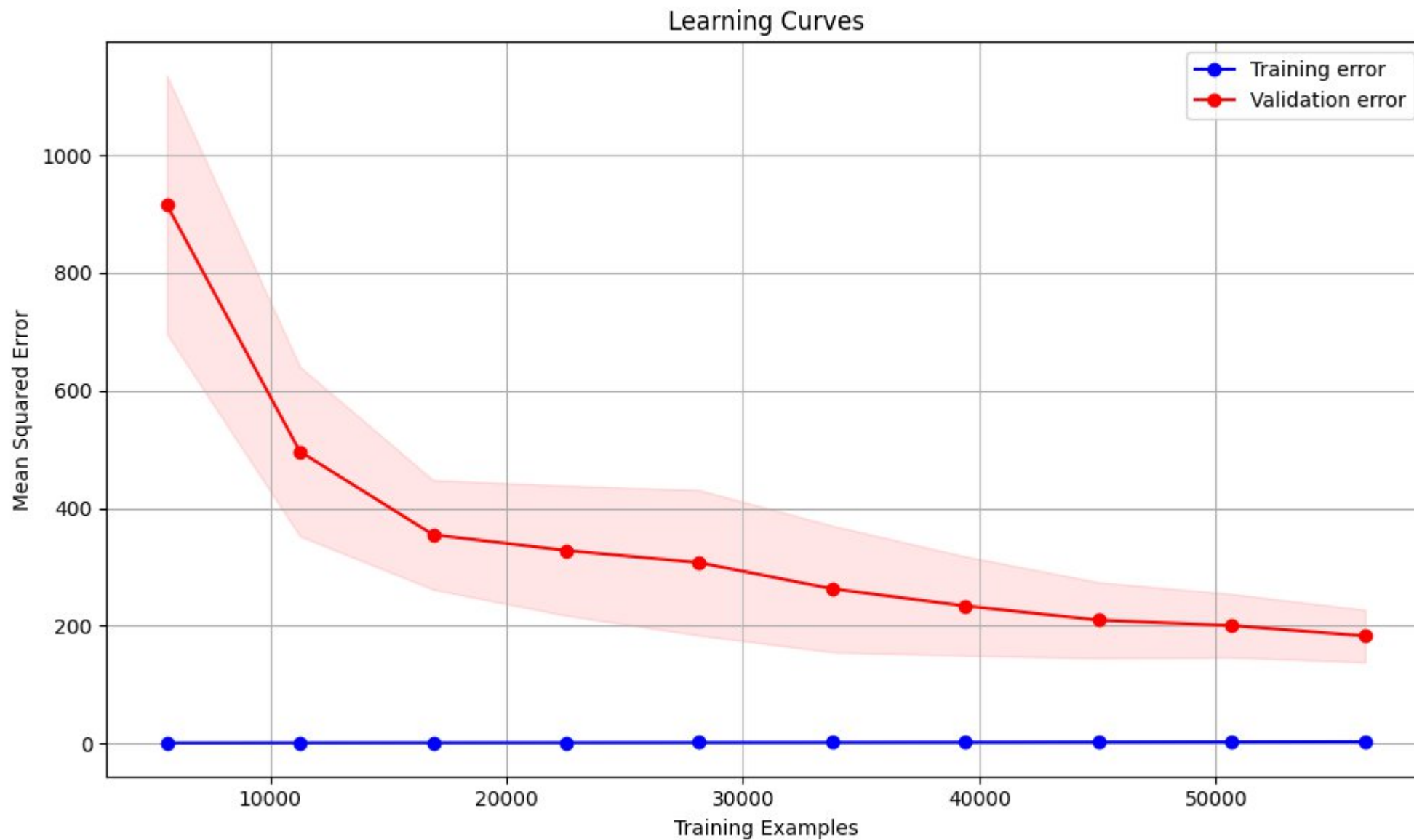
| | |
|----------------|-------|
| R2 for Winter: | 0.906 |
| R2 for Spring: | 0.825 |
| R2 for Summer: | 0.876 |
| R2 for Autumn: | 0.864 |



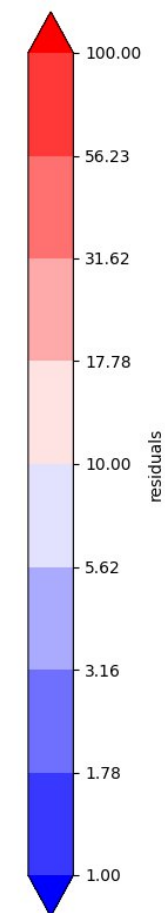
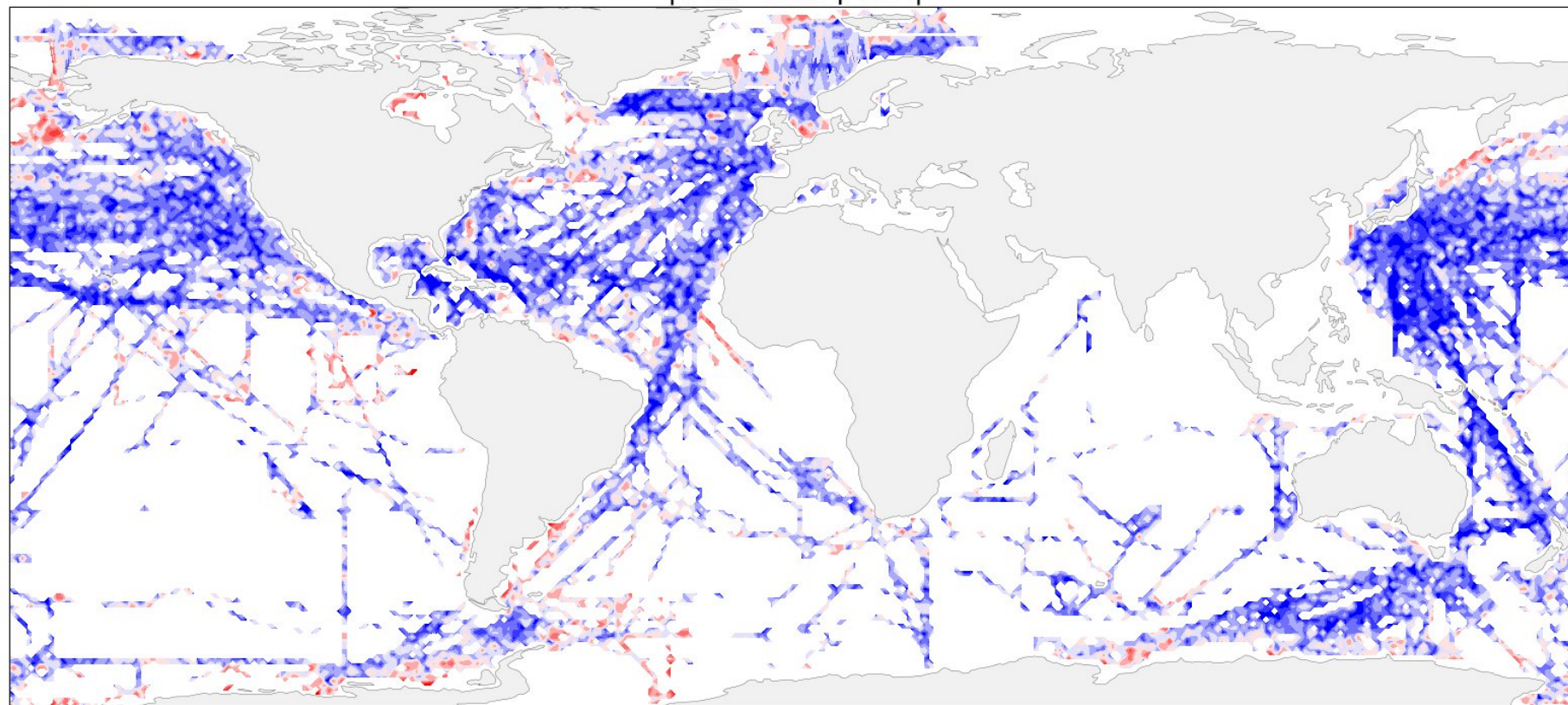
Residual Plotting



Learning Curve by Training Size



|True - Pred pCO₂|



Thank you for listening