Big Data Project : Ocean Warming

Carlos Freiji
Dhia Znaidi
Rayen Ben Ismail
Lucas Ye

Content

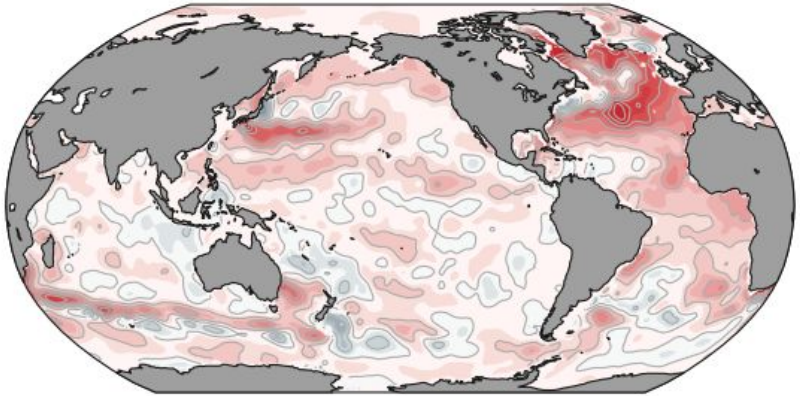
- Background
- Ocean Heat Content (OHC)
- Regression
- Conclusion

Overview

Because of the greenhouse effect, ocean is warming.

In this project, we calculate the Ocean Heat Content using the sea temperature recorded from 1950 to 2019.

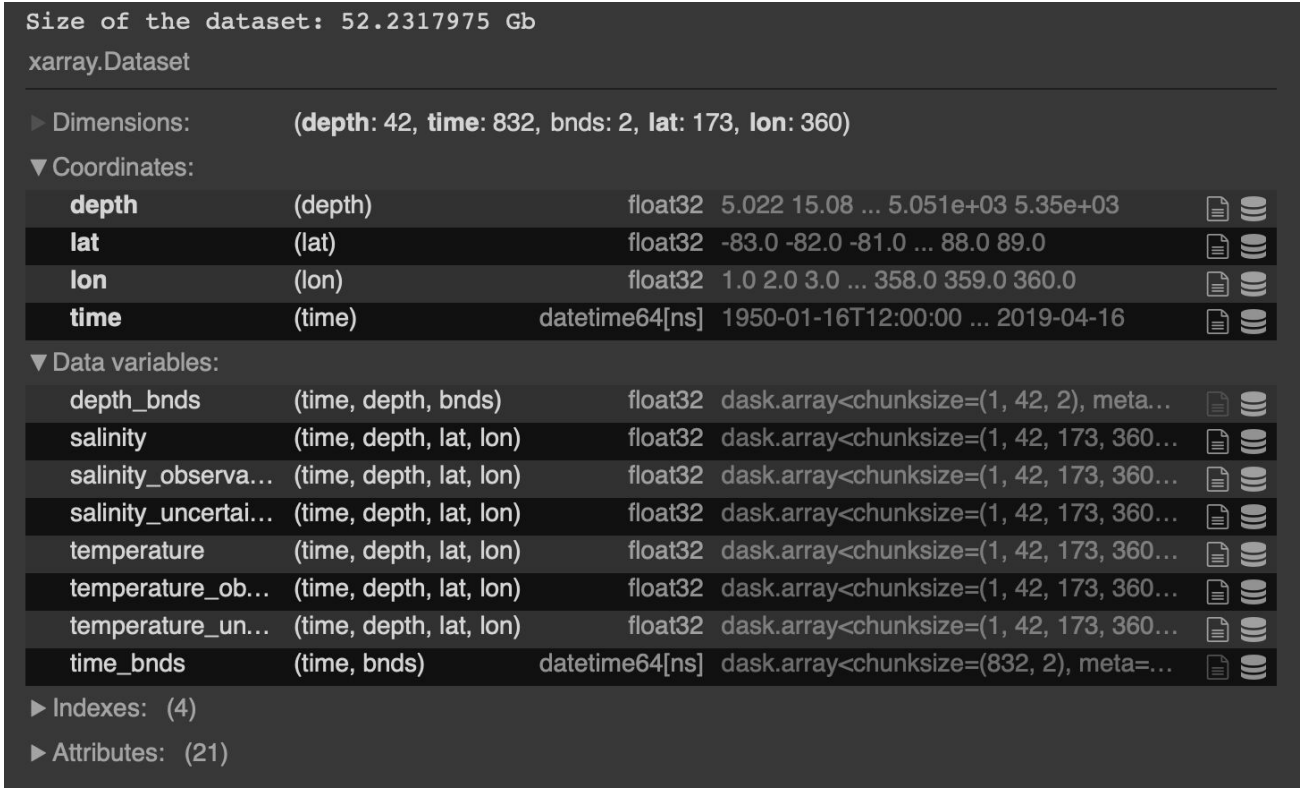
And then, we estimate the trend with the result of regression.



Background

- The warming of the oceans has accounted for approximately 93 % of the warming of the Earth since the 1950s. Warming of the upper (0–700 m) ocean accounted for about 64 % of the total heat uptake.
- A trend for increasing heat content in the upper ocean has become evident since the 1950s. Recent observations also show substantial warming of the deeper ocean (between depths of 700 and 2 000 m and below 3 000 m).
- Ocean heat content (OHC) is a measure of the amount of heat stored in the ocean. It is an important indicator of climate change because the ocean absorbs more than 90% of the heat added to the Earth's system by human activities. An increase in OHC can lead to changes in ocean circulation, sea level, and weather patterns, and can also have impacts on marine ecosystems.

Background - Dataset



Size of the dataset: 52.2317975 Gb

xarray.Dataset

▶ Dimensions: (**depth**: 42, **time**: 832, bnds: 2, **lat**: 173, **lon**: 360)

▼ Coordinates:

depth	(depth)	float32	5.022 15.08 ... 5.051e+03 5.35e+03
lat	(lat)	float32	-83.0 -82.0 -81.0 ... 88.0 89.0
lon	(lon)	float32	1.0 2.0 3.0 ... 358.0 359.0 360.0
time	(time)	datetime64[ns]	1950-01-16T12:00:00 ... 2019-04-16

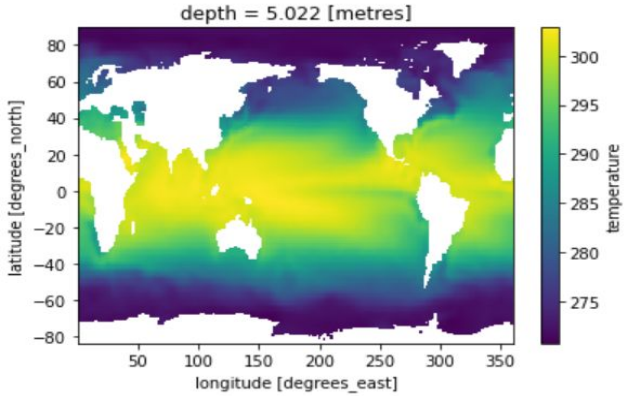
▼ Data variables:

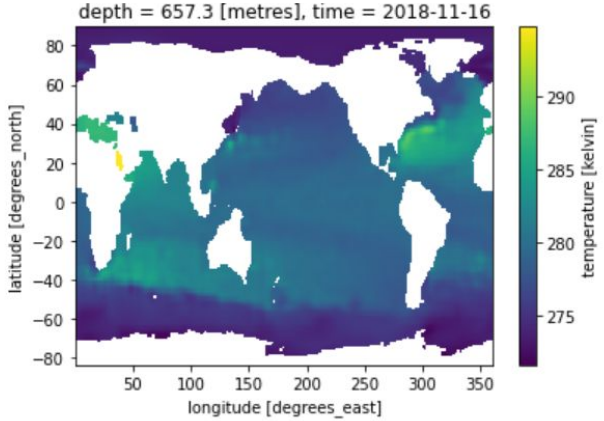
depth_bnds	(time, depth, bnds)	float32	dask.array<chunksize=(1, 42, 2), meta…
salinity	(time, depth, lat, lon)	float32	dask.array<chunksize=(1, 42, 173, 360…
salinity_observa…	(time, depth, lat, lon)	float32	dask.array<chunksize=(1, 42, 173, 360…
salinity_uncertai…	(time, depth, lat, lon)	float32	dask.array<chunksize=(1, 42, 173, 360…
temperature	(time, depth, lat, lon)	float32	dask.array<chunksize=(1, 42, 173, 360…
temperature_ob…	(time, depth, lat, lon)	float32	dask.array<chunksize=(1, 42, 173, 360…
temperature_un…	(time, depth, lat, lon)	float32	dask.array<chunksize=(1, 42, 173, 360…
time_bnds	(time, bnds)	datetime64[ns]	dask.array<chunksize=(832, 2), meta=…

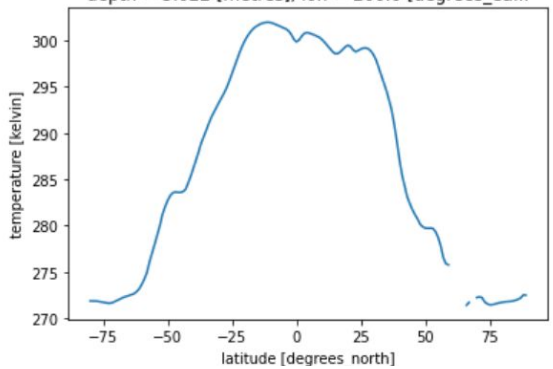
▶ Indexes: (4)

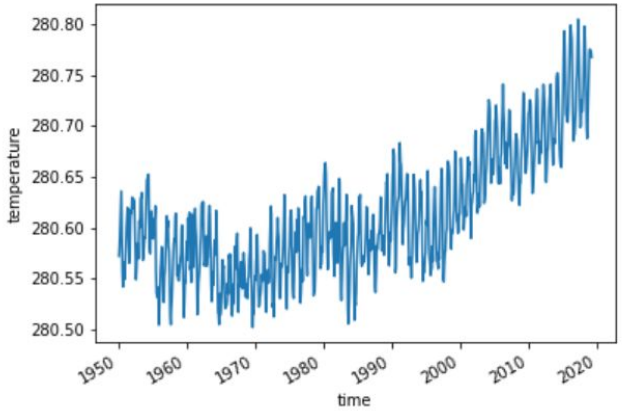
▶ Attributes: (21)

Background - Dataset

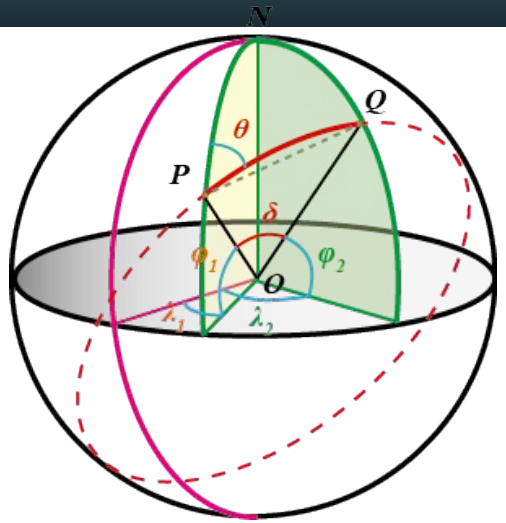








Ocean Heat Content (OHC)



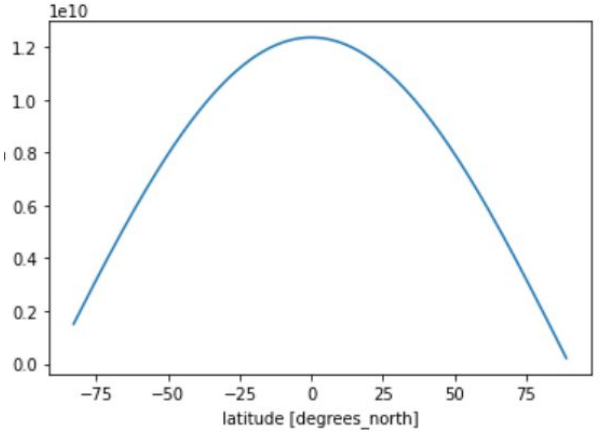
$$OHC(x,y,z,t) = \rho \cdot Cp \cdot dV(x,y,z) \cdot T(x,y,z,t)$$

$$dV(x,y,z) = dA(x,y) \cdot dH(z)$$

$$dA(x,y) = R^2 \cdot d\phi \cdot d\lambda \cdot \cos(lat[rad])$$

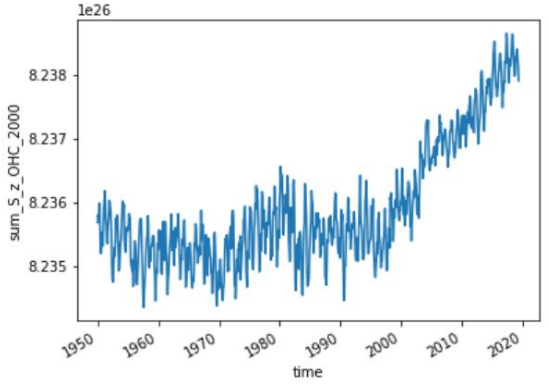
$$d\phi = d\lambda = np.deg2rad(1.)$$

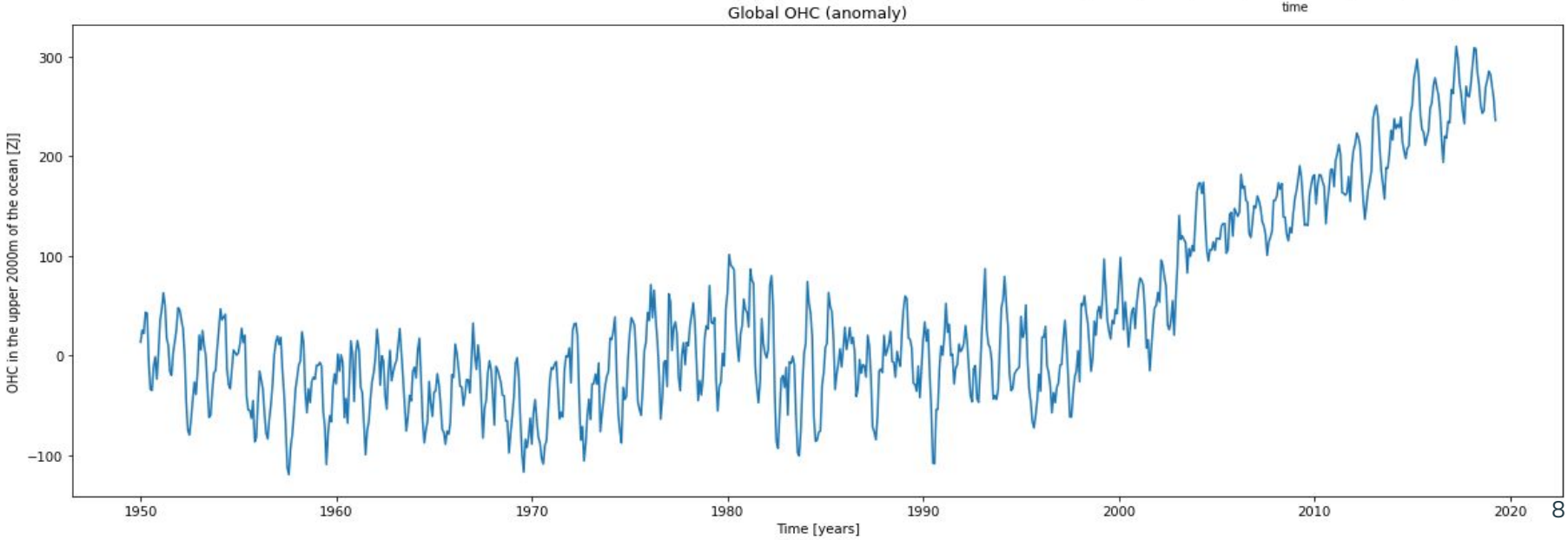
- ρ = 1035 kgm^{-3}
- C_ρ = 4,186 J/g°C



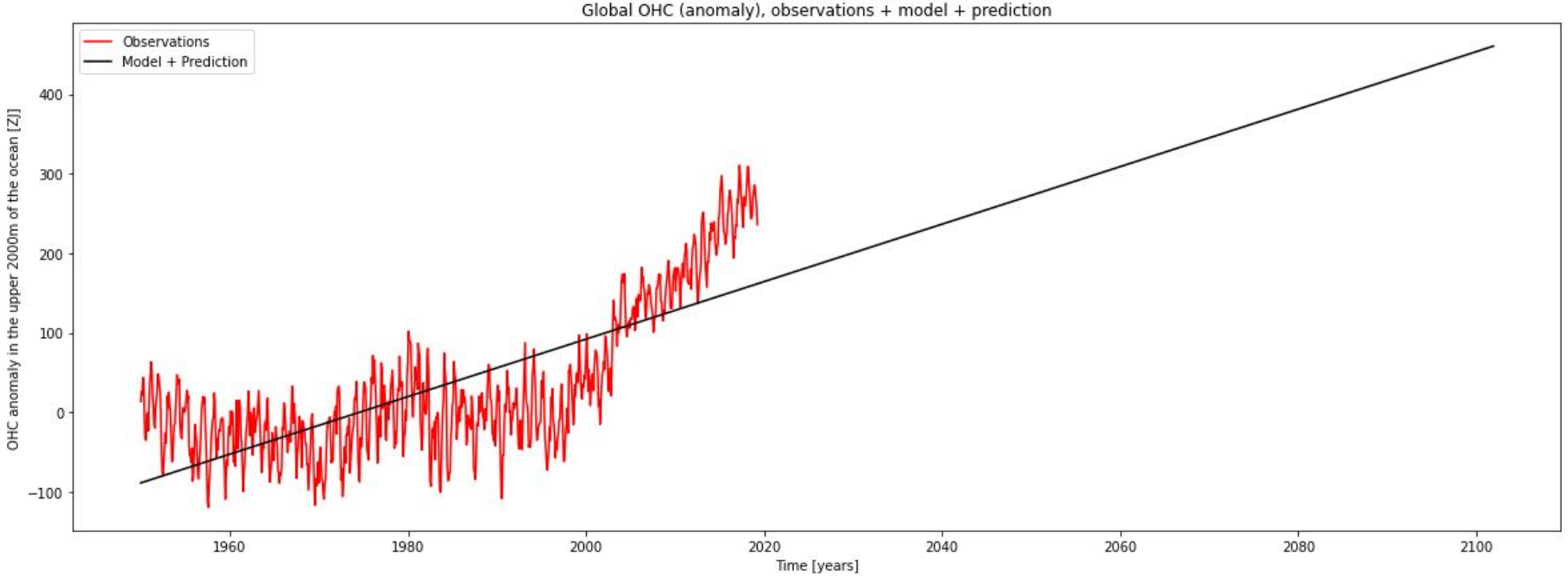
Ocean Heat Content (OHC)

Global OHC (1950-2019)

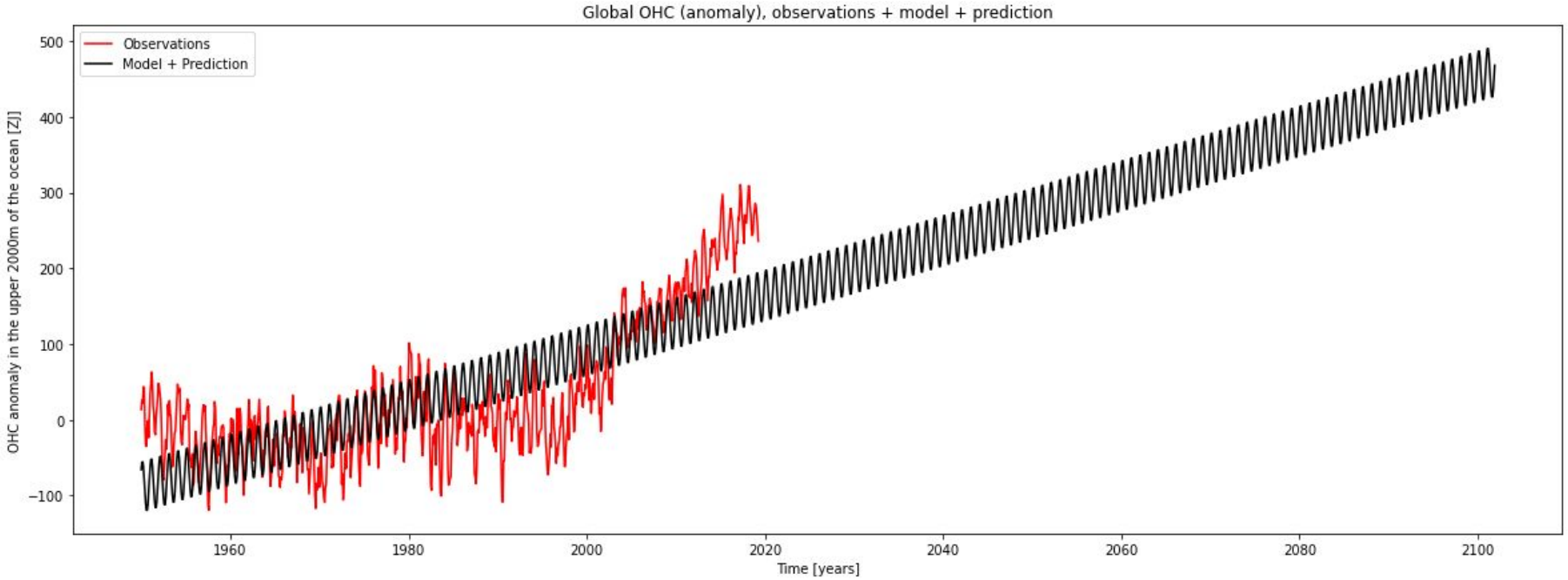




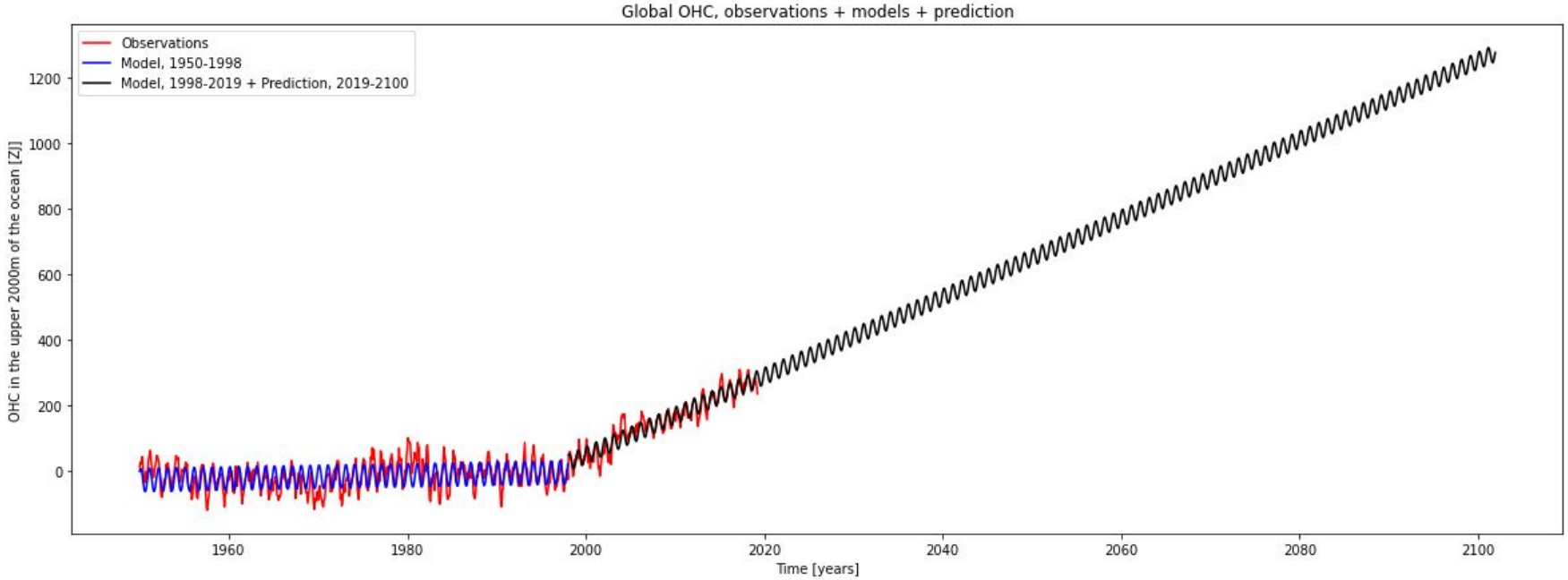
Regression Models : Linear regression



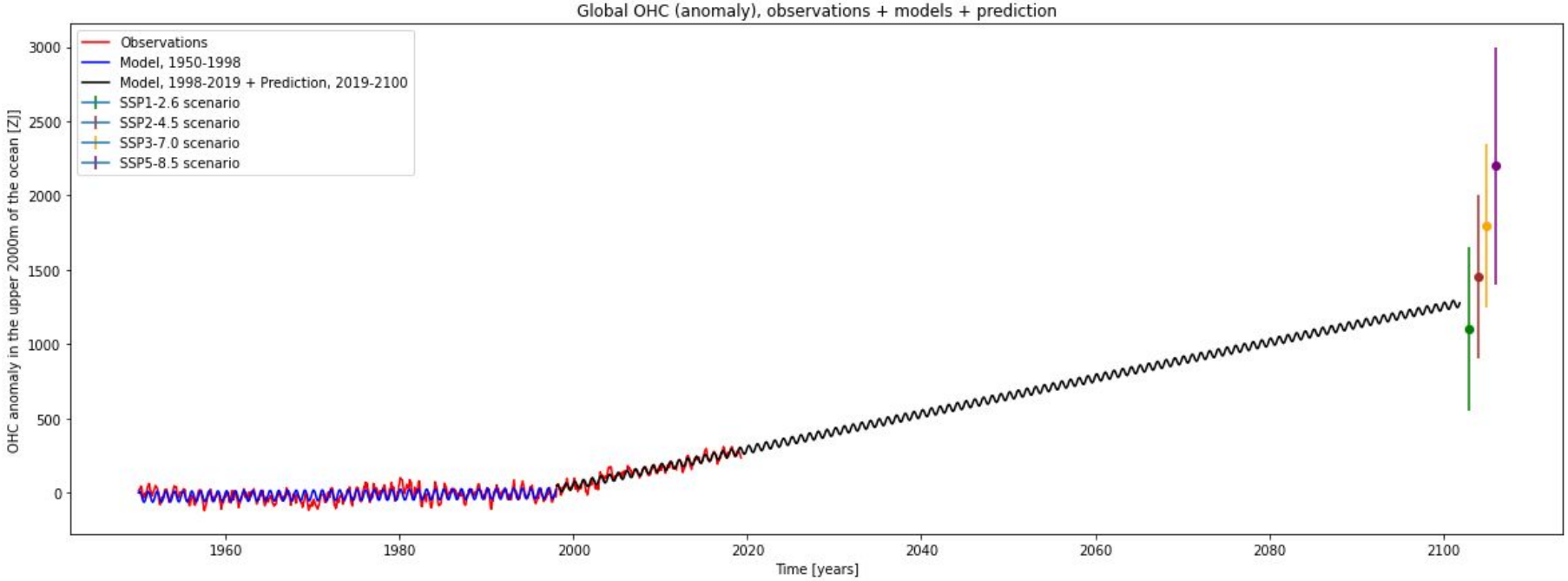
Adding new features : cos and sin (seasons)



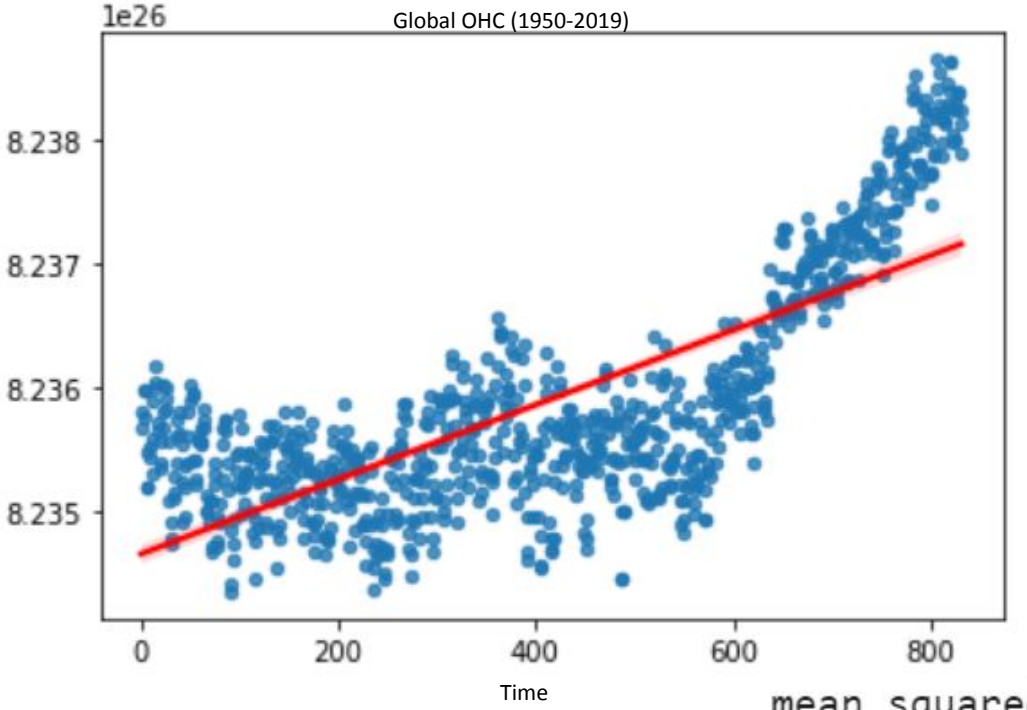
Linear regression - dataset separation



Linear regression - dataset separation
Comparison with IPCC predictions

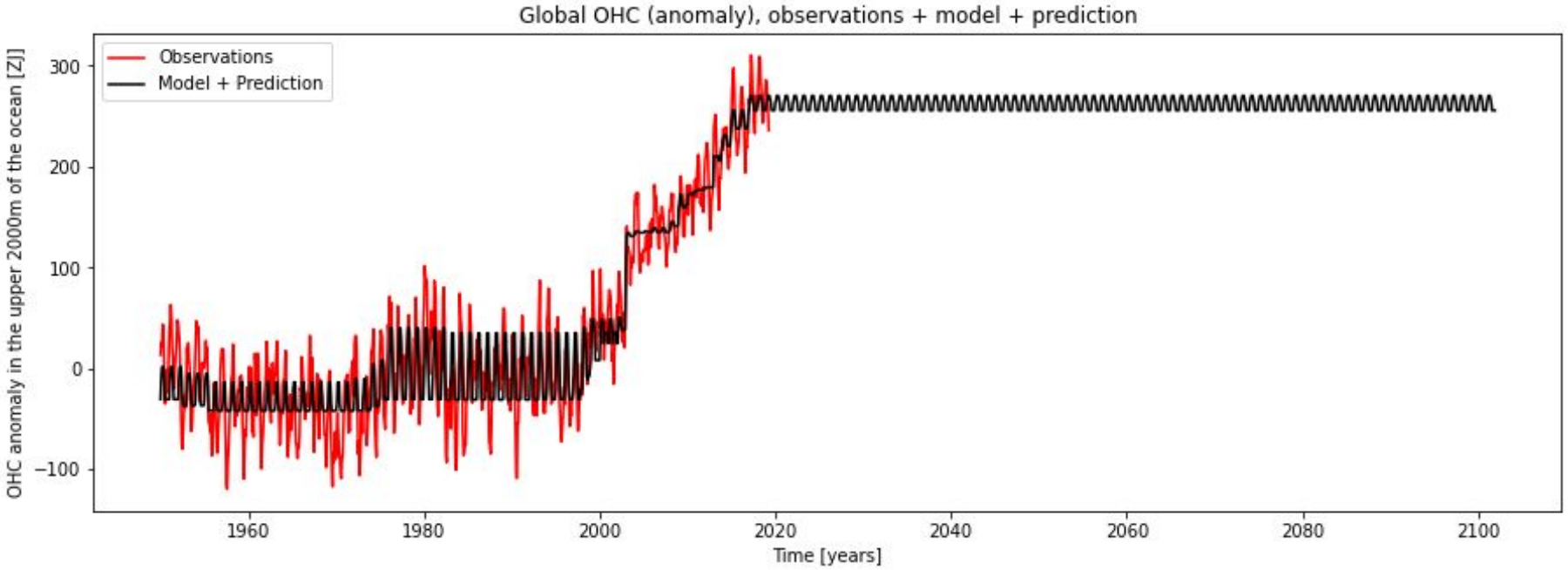


Confidence intervals

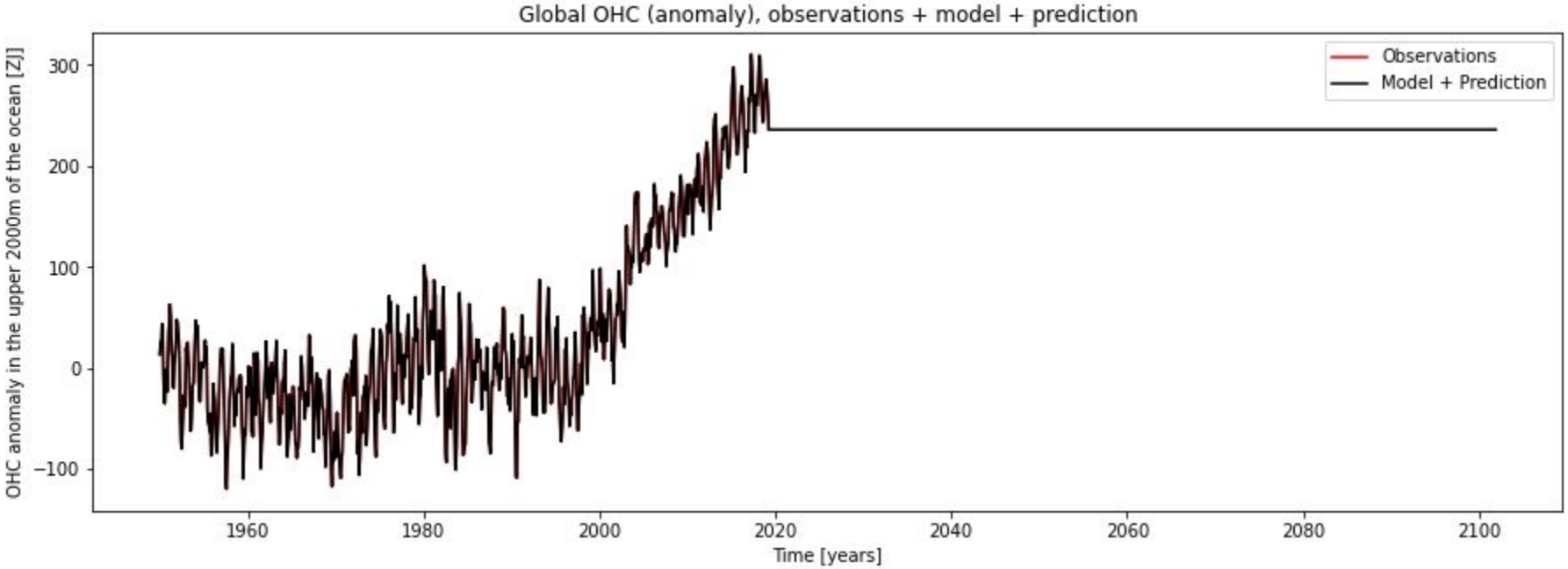



```
mean squared error= 3.9426716282250 94e+45
mean absolute error =  23787511177.846153
```

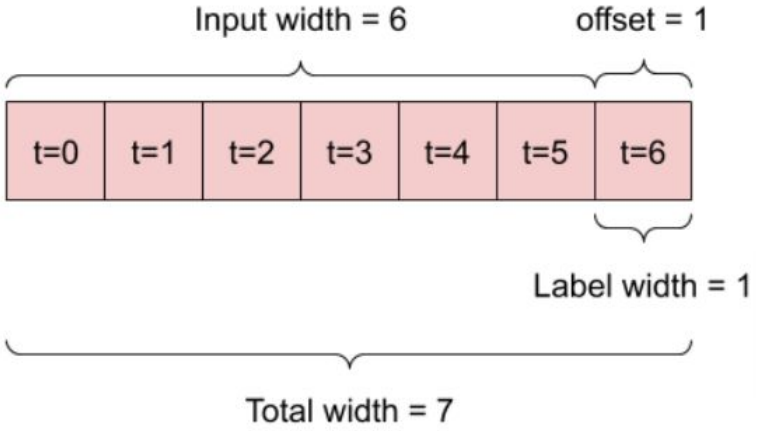
Adaboost regressor

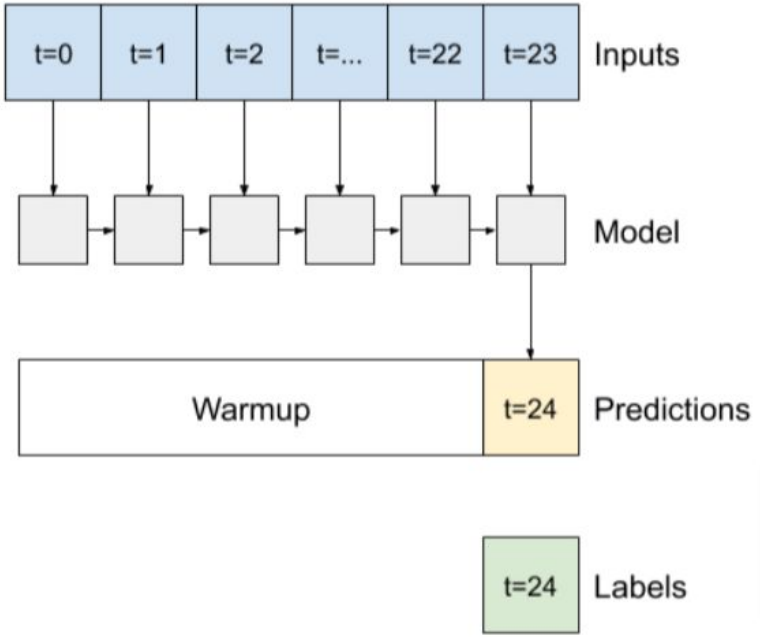


Random forest regressor

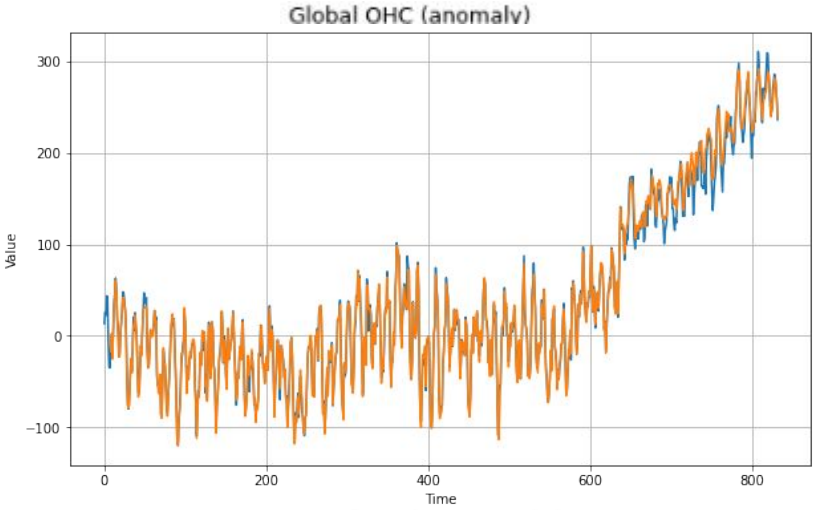


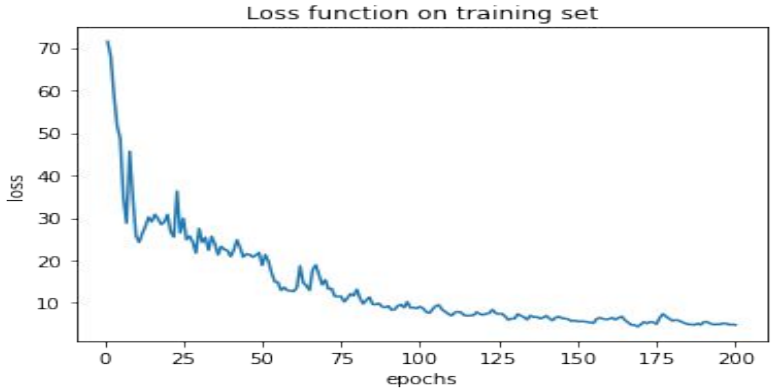
RNN Model

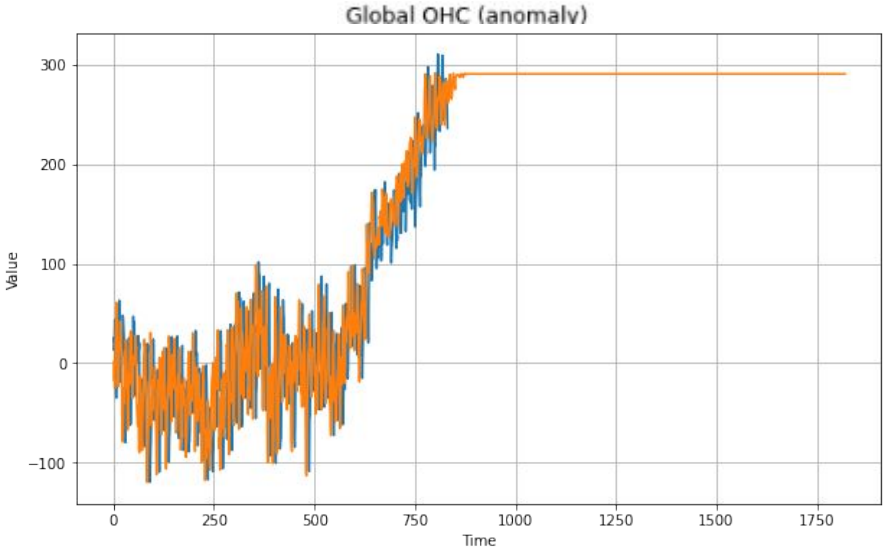

Input width = 6
offset = 1
t=0
t=1
t=2
t=3
t=4
t=5
t=6
Label width = 1
Total width = 7


t=0
t=1
t=2
t=...
t=22
t=23
Inputs
Model
Warmup
t=24
Predictions
t=24
Labels

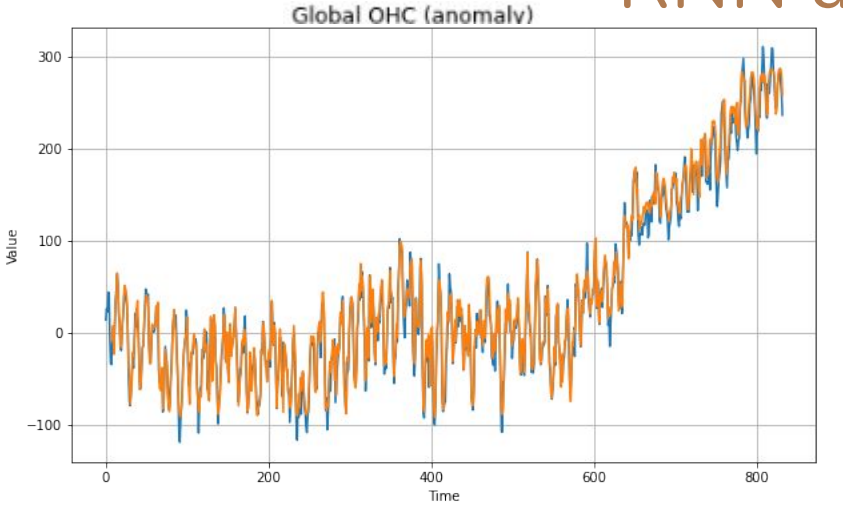
RNN approach

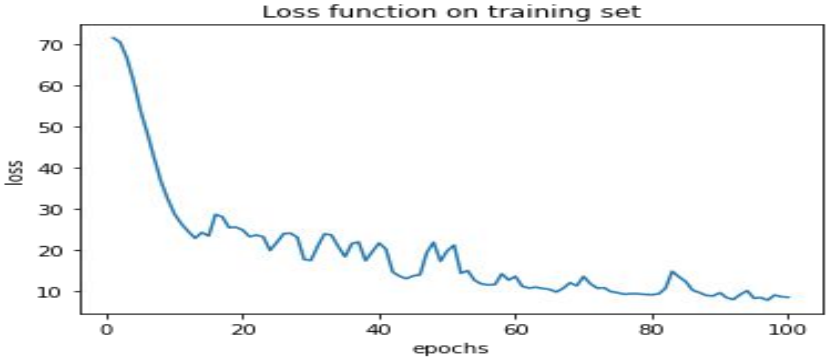


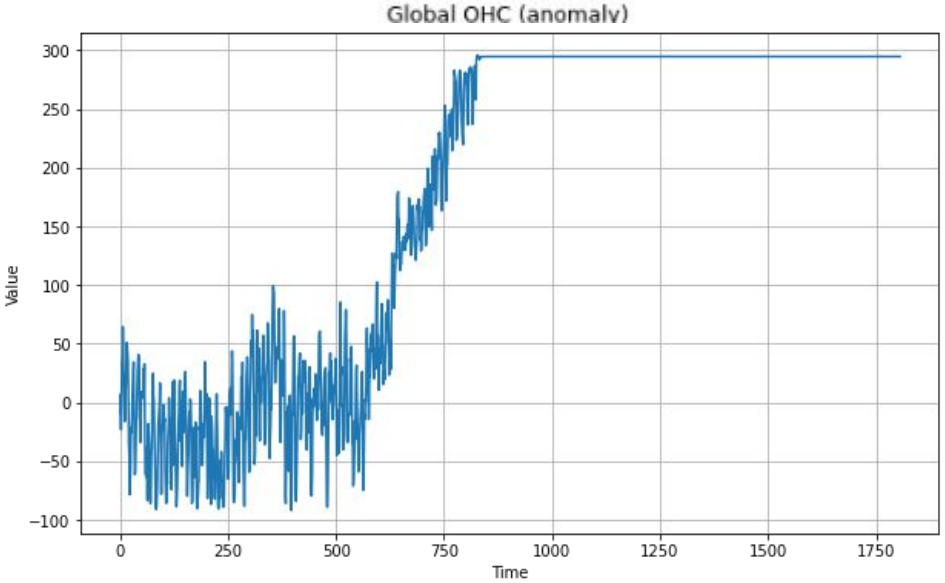




RNN approach

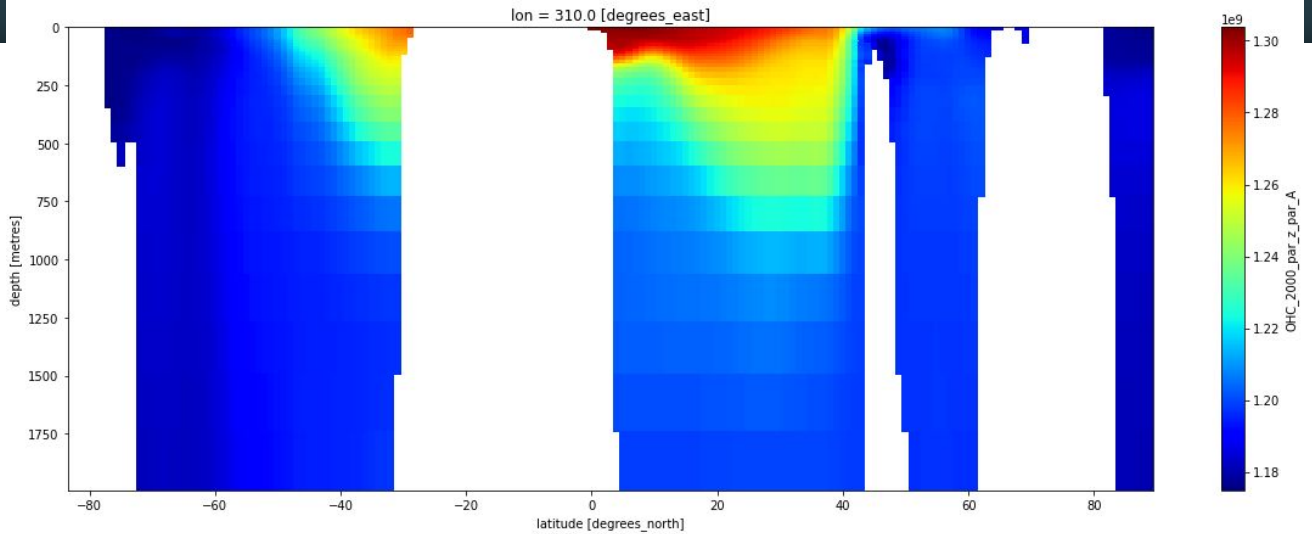




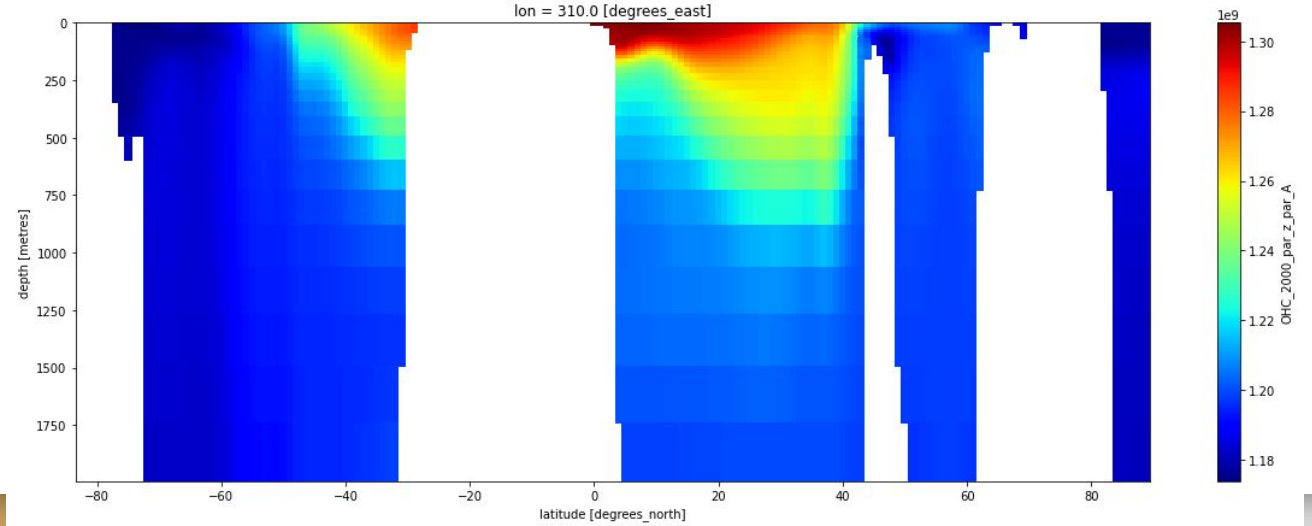


Locally

Mean OHC, for the years 1950-58, at lon=50W

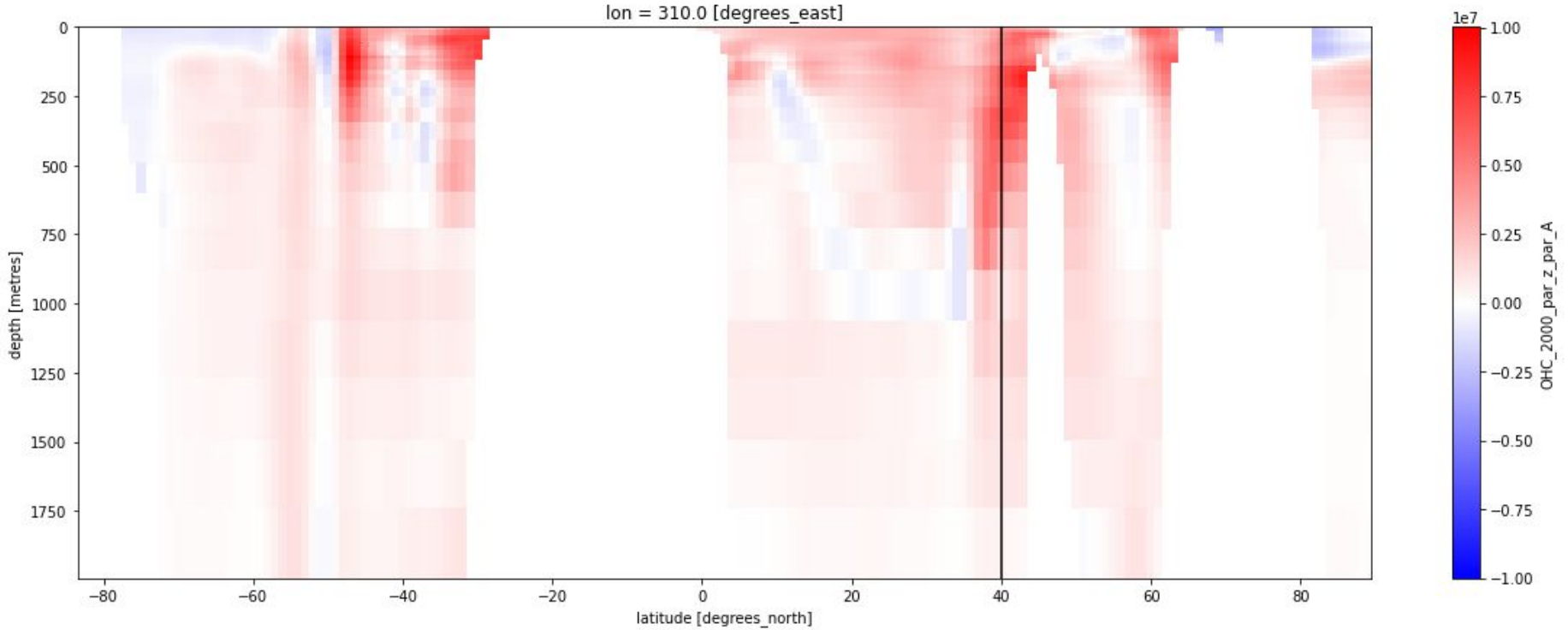


Mean OHC, for the years 2010-18, at lon=50W



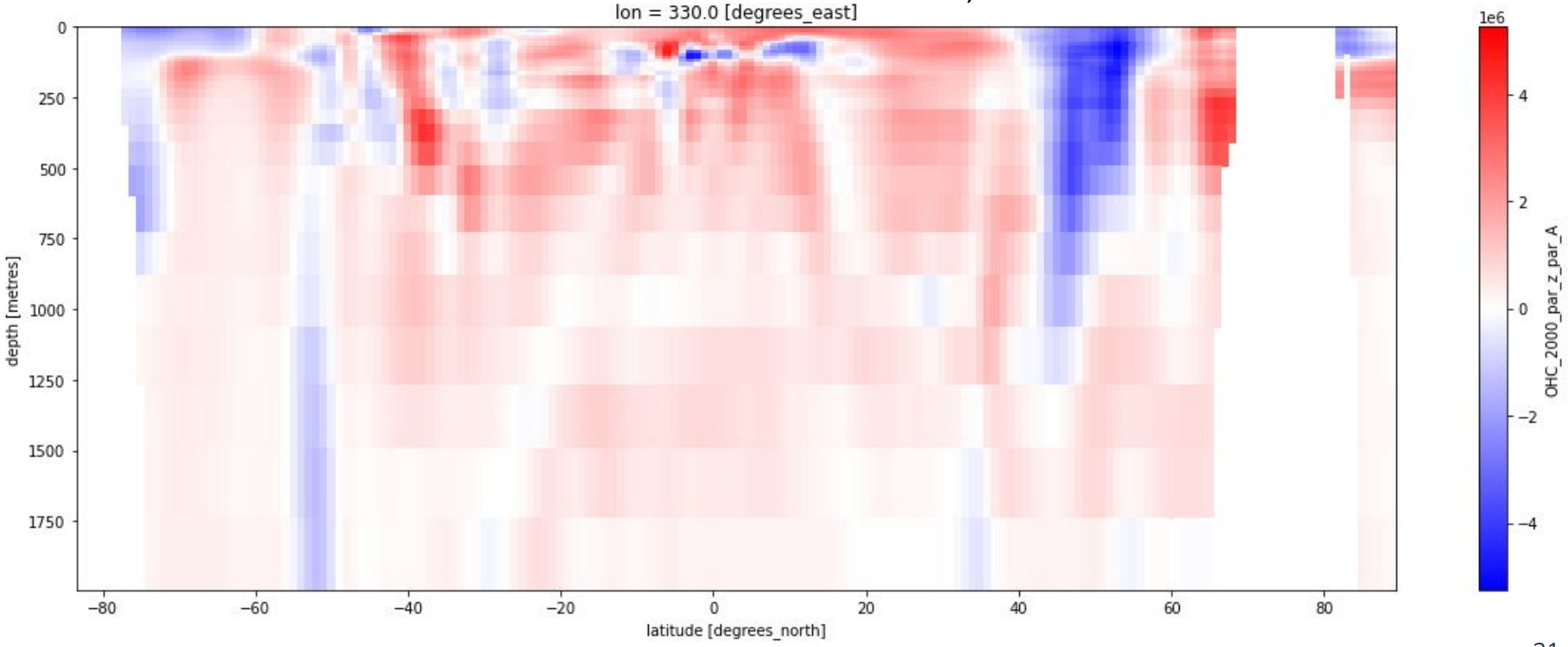
Locally

OHC difference between 2010-18 and 1950-58, at lon=50W



Locally

OHC difference between 2010-18 and 1950-58, at lon=30W



Conclusion

Linear Regression :

- Global trend seems to be well represented
- However, not very complex : local variations

Other models didn’t produce results and forecasts as expected.

The problem may be as follows:

1. Lack of sufficient information
2. Overfitting over the training examples
3. Too complex models for the forecasting task