Analyzing Sea Surface Temperature Using Autoencoders

Louis Kamp Eskildsen - S203977   (DTU compute)

1. SST and The Climate

Oceans cover 70% of the Earths surface and is responsible for driving many of the vital parts of the climate. The temperature of the Oceans can have a major effect on local climate as seen by the Gulf Stream that lifts and regulates the temperature in western Europe by bringing hot water from the Gulf of Mexico. Another example of effects caused by sea surface temperature (SST) is El niño. El niño is a phenomenon located in the Pacific out from the coast of South America where hot water rise to the surface. When El niño occurs it is known to cause higher than normal temperatures around the world. El niño has also been linked to droughts and extreme rainfall. Therefore, predicting when an El niño cycle begins and ends helps in predicting the global climate. In this project it has been shown that the global SST can partially be encoded by an autoencoder and forecasted using a DMD model.
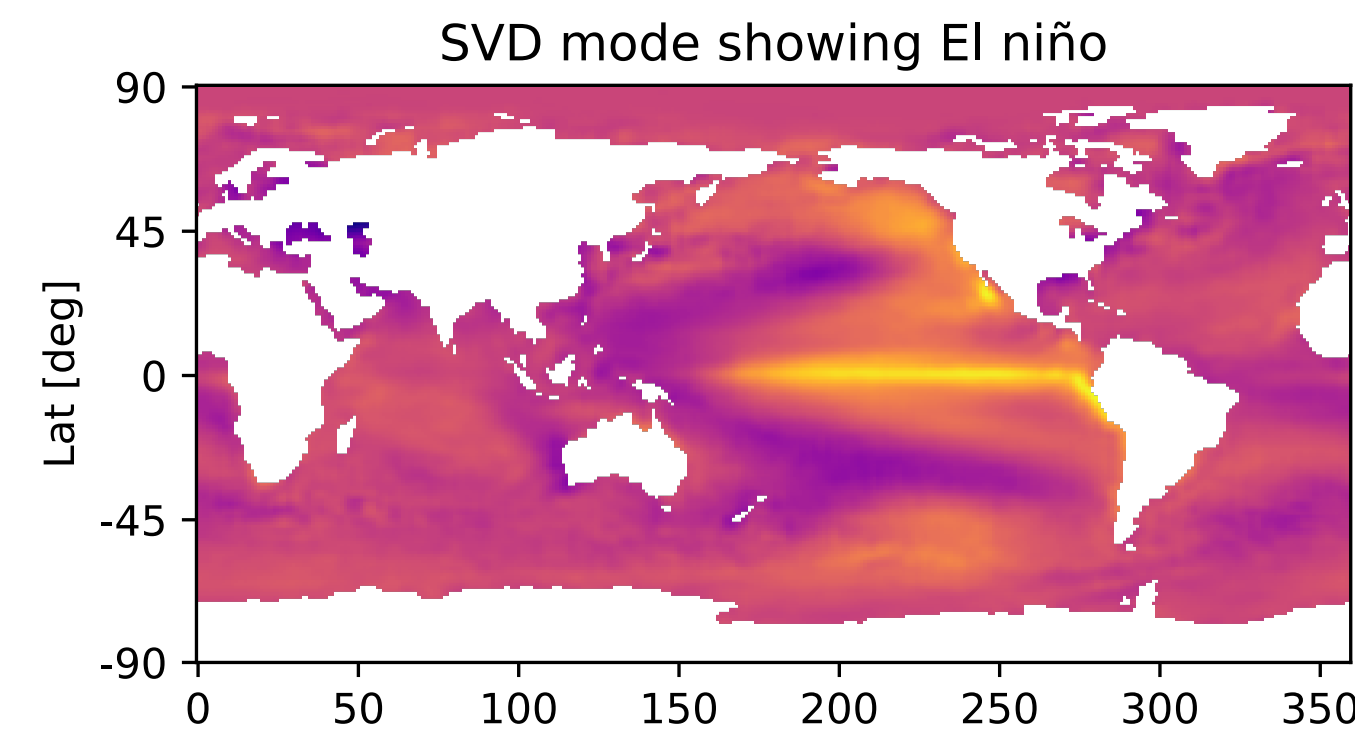

Figure 1: This plot shows the El niño as a mode of the singular value decomposition.

Previous studies have shown that a partial prediction of SST can be done with dynamic mode decomposition (DMD) [1]. This project shows that the DMD algorithm can be used in combination with an autoencoder to simplify computational complexity by reducing the number of parameters to model. The SST dataset used in this project comes from NOAA Optimum Interpolation and gives the average monthly SST from 1981 to 2022 with 44.219 datapoints each month [2].
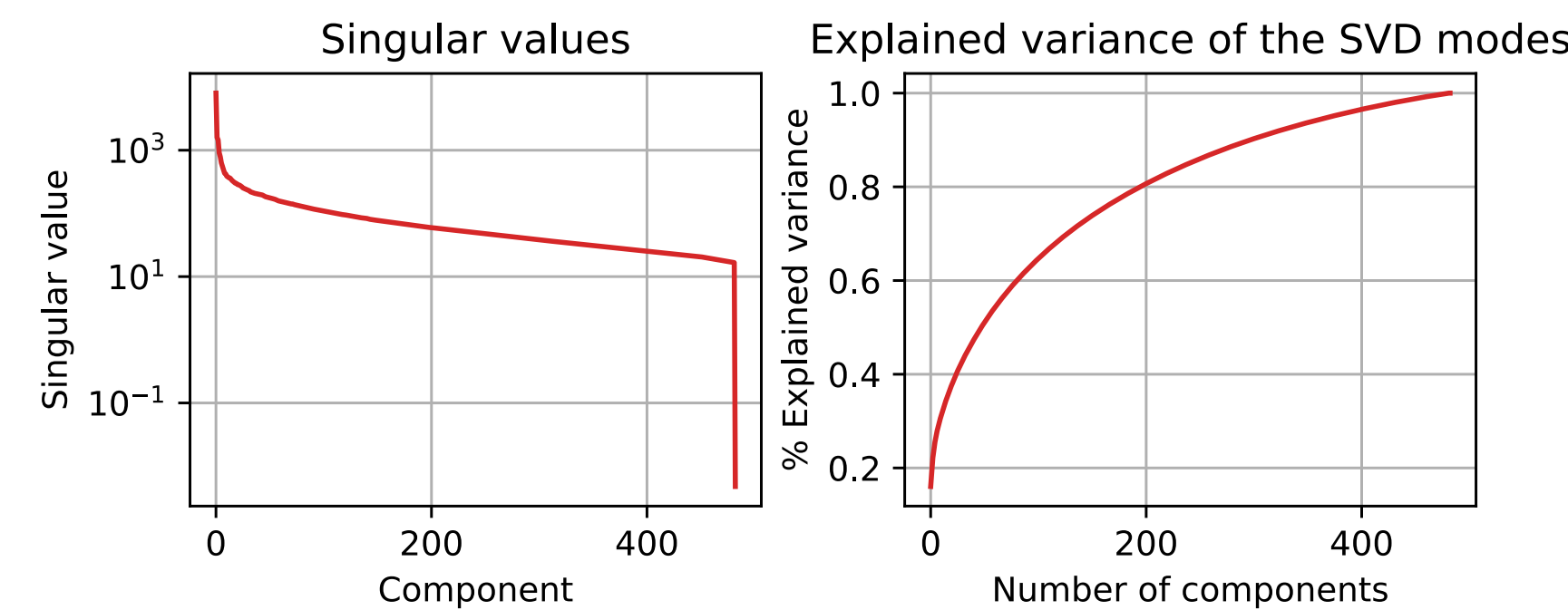

Figure 2: This figure show the SVD singular values and the cumulative explained variance of the SVD modes.

There are multiple ways of encoding SST in a latent space. A commonly used method is the singular value decomposition (SVD). To estimate the size of the latent space of the autoencoder, an SVD analysis is carried out. Figure 2 shows the power of each sigular value and the cumulative explained variance captured by the SVD modes. The third of these modes can be seen in Figure 1 and shows the carceraristics of EL niño phenomenon. The power of each singular value is seen to be decreasing before leveling off at approximately the 500'th component. This tells us that the underlying dimensionality of the dataset is around 500. Basing the autoencoder on a latent size of this is therefore a good starting point.

[1] Eduardo Rodrigues, Bianca Zadrozny, Campbell Watson, and David Gold, "Decadal Forecasts with Res-DMD: a Residual DMD Neural Network." Jun. 21, 2021.

[2] "NOAA Optimum Interpolation (OI) SST V2." [Online]. Available: https://www.psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html

2. Autoencoder for SST

To encode the SST an autoencoder based on a feedforward neural network has been implemented in PyTorch. The autoencoder will be evaluated with the mean squared error. Let the operator $\mathcal{N}$ represent the transformation from the true SST to the latent space by the neural network . Then latent mapping is then given by: $y_i = \mathcal{N}x_i$ and the inverse by $x_i \approx \mathcal{N}^{-1}y_i$. By minimizing the MSE on $x_i \approx \mathcal{N}^{-1}\mathcal{N}x_i$, the neural network learns how encode and decode the SST dataset.

The advantages of using an autoencoder over the latent space of the SVD is that the autoencoder might give a better representation of the non-linear sates since the feedforward neural network is inherently non-linear due to its activation functions.

2.a. The architecture of the model

The autoencoder is composed of two large hidden layers as well as one layer in the middel that represents the latent space. ReLu is used as the activation function between the input and hidden layer as well as the hidden and output layer. To prevent overfitting a dropout layer is placed between the input and hidden layer during training.
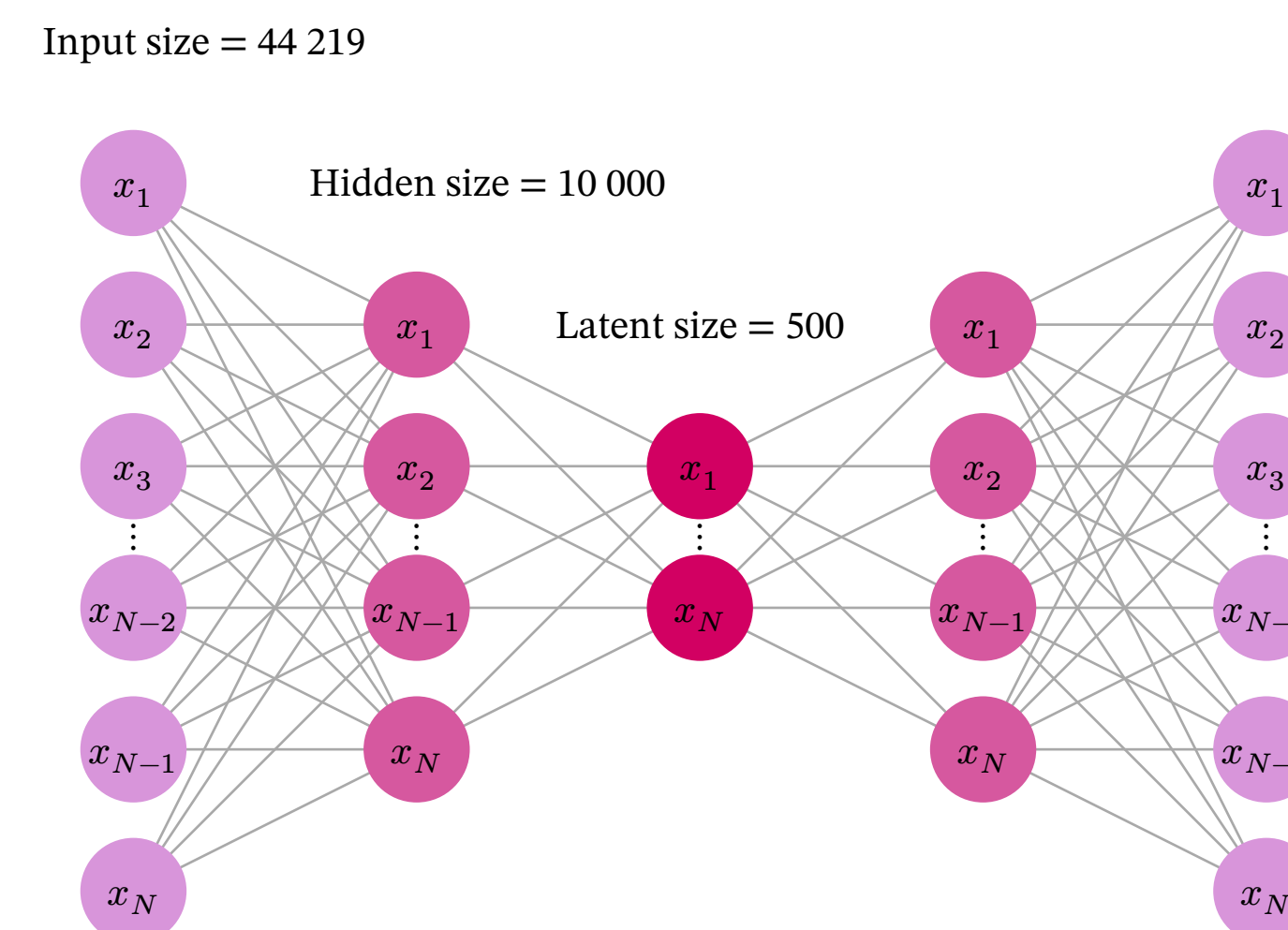

Figure 3: The architecture of the autoencoder network.

2.b. Training and Tuning With Validation Data

The SST dataset is split into three partisions: 70%: training, 20%: validation, 10%: testing.

The process of training the autoencoder will in this project happen in two stages. (1) Initial training: In this step the training dataset will be used to train the encoding and decoding of the SST into the latent space. (2) Tuning: This step will use the validation dataset as input to optimize the models ability to encode, forecast and decode the data. It was found that the Adam optimizer worked best for initial training and SGD optimizer for the tuning.
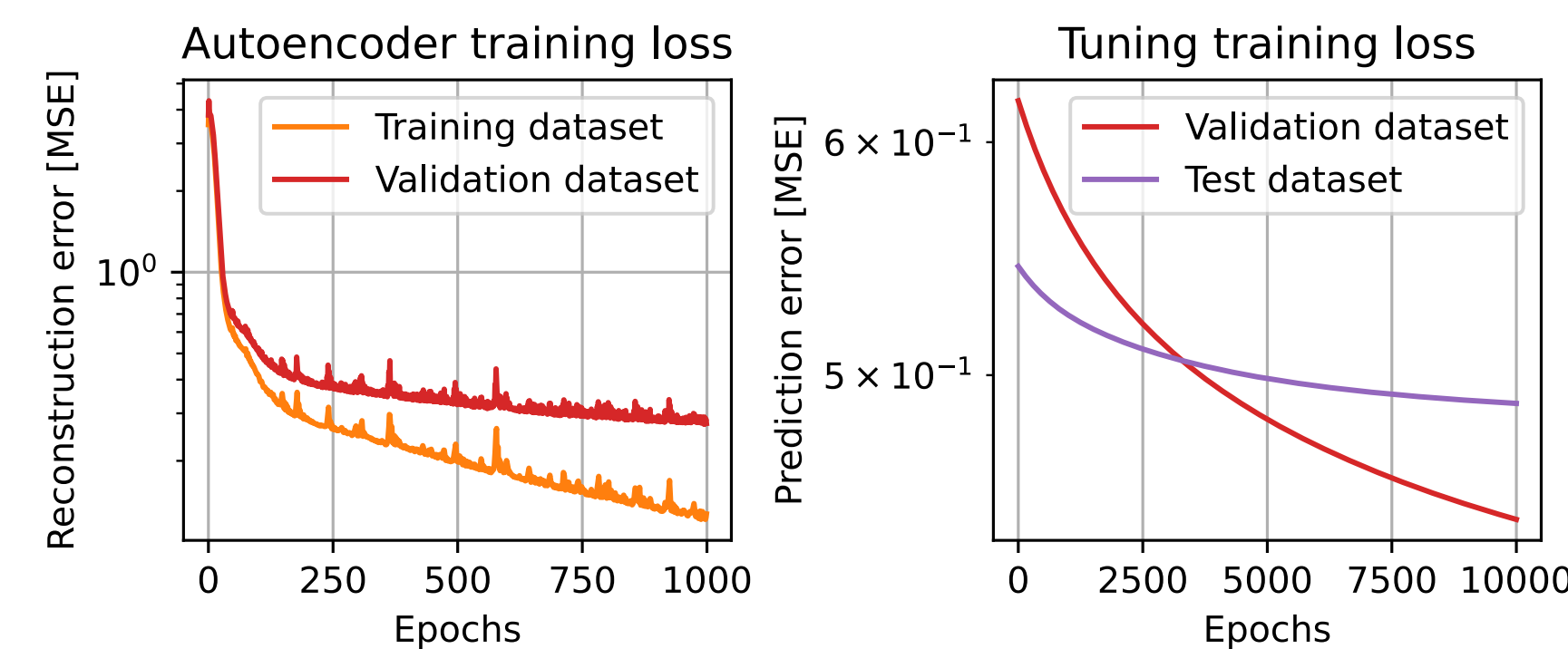

Figure 4: (Training): The final train loss: 0.13, validation loss: 0.28.
(Tuning): The final validation loss: 0.44, test loss: 0.50.

3. Forecasting SST in latent space with DMD

To forecast the SST the DMD method will be used. In order to use the DMD method the encoded training data snapshots are placed into two matrices:

$$W' = [y_i], i = 2...n, \quad \text{and} \quad W = [y_i], i = 1...(n-1). \tag{1}$$

The DMD matrix A is found by using the pseudoinverse of W:

$$W' = AW \rightarrow A = W'W^{\dagger} \tag{2}$$

By solving for A a linear model between one snapshot and the next has been made. To predict the SST in the latent space the following model can be used: $x_i \approx \mathcal{N}^{-1}A^i\mathcal{N}x_0$. The figures below show the forecasted SST, which are predicted from February 2018, the first snapshot of the testing dataset. The prediction map shows an El niño phenomenon happening in September of 2021, and the error map shows this prediction to be wrong since the real El niño phenomenon happened in 2019. Comparing the model RMSE map and SST STD map also shows that the model has trouble in predicting where SST varies.
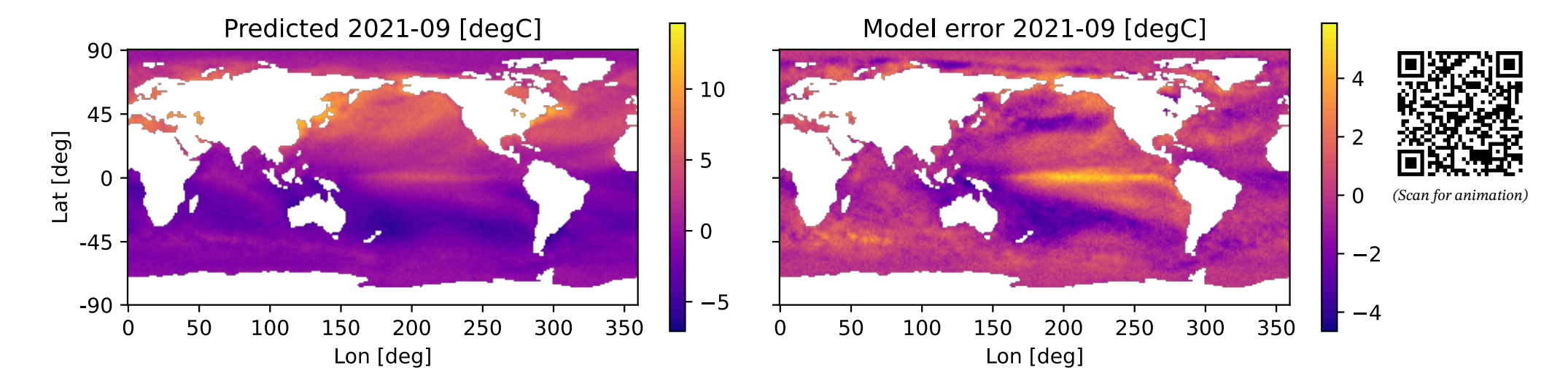

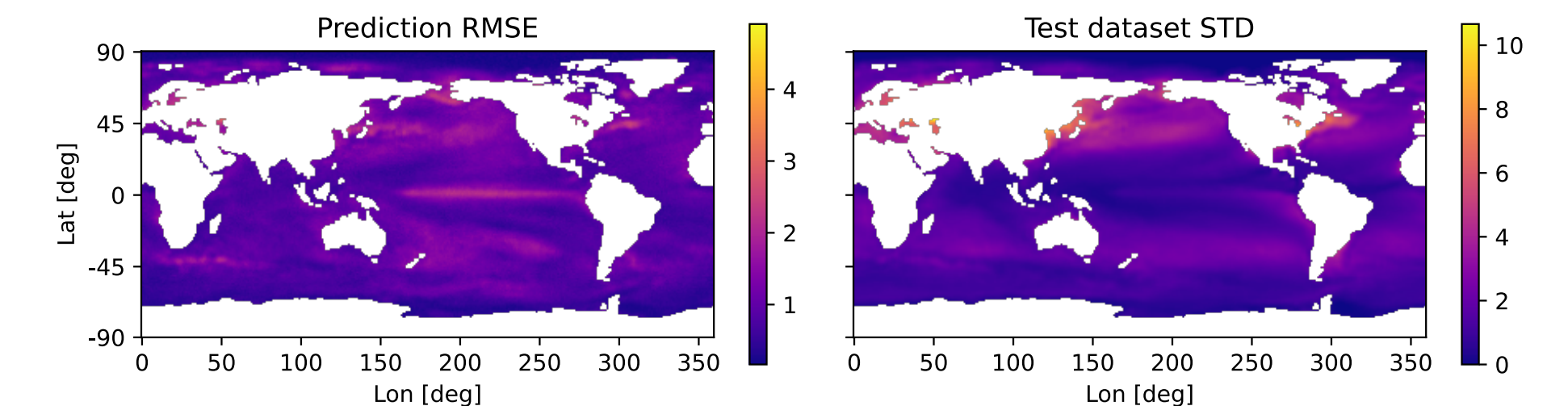Figure 5: This figure show how the models predicts an El niño event in Marts of 2020.

Figure 6: This figure show that the RMSE is highly correlated with the STD of the SST.

4. Conclusion

These results suggest that an autoencoder is capable of encoding the SST data into a latent space, and the DMD algorithm is able to pick up on some of the dynamics of the system but is not capable of accurately long term forecasting. Figure 7 shows that the performance of the autoencoder is slightly worse than using the latent space of the SVD in combination with DMD. The predictions might be improved by replacing DMD with a more flexible method such as a RNN or LSTM architecture.
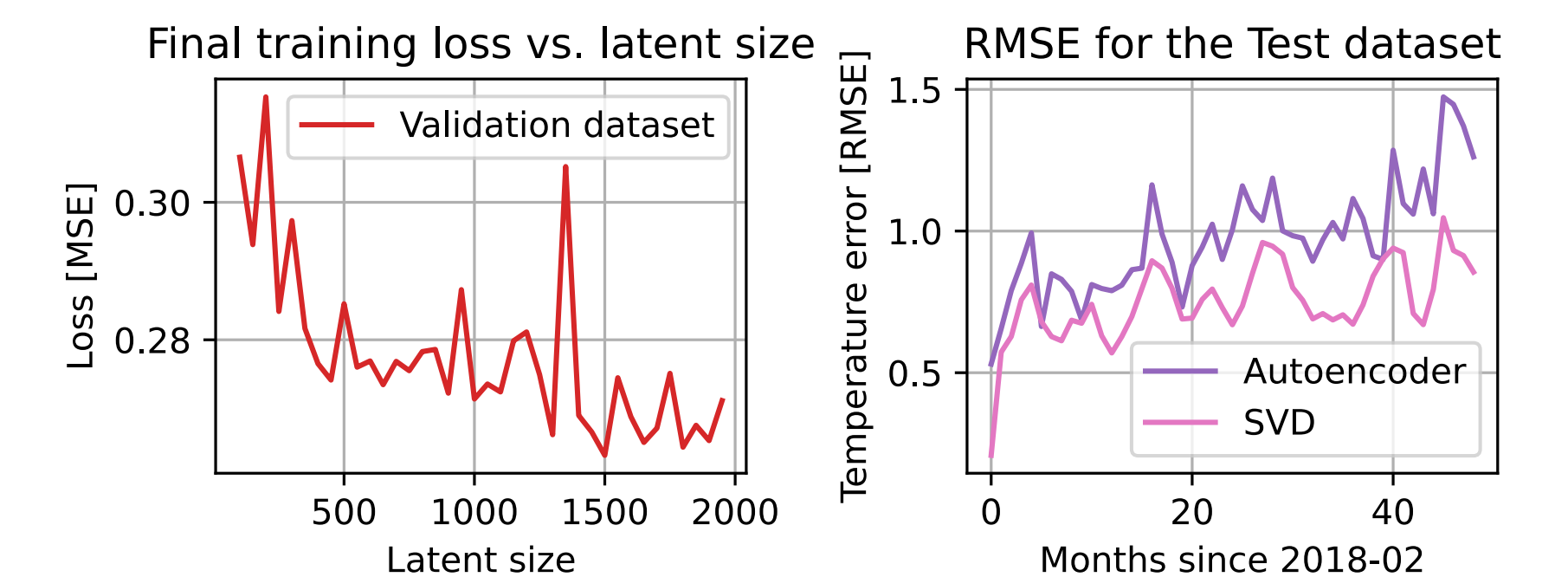

Figure 7: (Left): The plot shows the final loss for a model with a given latent space. The plots confirms that the optimal latent space is around 500. (Right): The mean squared error of the model as it predicts the SST from the first snapshot in the test dataset.