

---

# Decadal Forecasts with ResDMD: a Residual DMD Neural Network

---

Eduardo Rodrigues<sup>1</sup> Bianca Zadrozny<sup>1</sup> Campbell Watson<sup>1</sup> David Gold<sup>2</sup>

## Abstract

Operational forecasting centers are investing in decadal (1-10 year) forecast systems to support long-term decision making for a more climate-resilient society. One method that has previously been employed is the Dynamic Mode Decomposition (DMD) algorithm – also known as the Linear Inverse Model – which fits linear dynamical models to data. While the DMD usually approximates non-linear terms in the true dynamics as a linear system with random noise, we investigate an extension to the DMD that explicitly represents the non-linear terms as a neural network. Our weight initialization allows the network to produce sensible results before training and then improve the prediction after training as data becomes available. In this short paper, we evaluate the proposed architecture for simulating global sea surface temperatures and compare the results with the standard DMD and seasonal forecasts produced by the state-of-the-art dynamical model, CFSv2.

## 1. Introduction

In recent years there has been significant effort by major dynamical modeling groups to perform decadal (1-10 year) forecasts (e.g., [Yeager et al. \(2018\)](#) and [Vecchi \(2015\)](#)). These forecasts can bridge the gap from seasonal forecasts ([Kirtman et al., 2014](#)) to multidecadal predictions ([Eyring et al., 2016](#)) and are specifically designed to forecast fluctuations through knowledge of the current climate state and multi-year ocean variability. Although skill is limited by the predictability of phenomena like the El NiñoSouthern Oscillation (ENSO), decadal forecasts can provide critical and actionable information on regional climate trends.

Another strategy for decadal predictions that has received attention uses data to fit a statistical model. A particular example of such an approach is found in ([Newman, 2013](#)), in which the Linear Inverse Model, also known as Dynamic

Mode Decomposition (DMD), is applied to make decadal forecasts. DMD has been used to successfully model climatic phenomena such as tropical diabatic heating and sea surface temperatures (SSTs; [Huddart et al. \(2017\)](#)). In its most basic form, DMD seeks to fit a linear dynamical system of the form:

$$\frac{d}{dt}\mathbf{x} = \mathbf{Ax} \quad (1)$$

which has an exact solution given by:

$$\mathbf{x}(t_0 + t) = e^{\mathbf{A}t}\mathbf{x}(t_0) \quad (2)$$

The dynamics are characterized by the eigenvalues and eigenvectors of  $\mathbf{A}$ , typically the leading eigenvalue/eigenvector pairs.

A common strategy in modeling environmental variables with DMD is to assume that the non-linear dynamics can be approximated by a linear system with random noise ([Newman, 2013](#); [Shin et al., 2020](#)), usually Gaussian. In this paper, we seek a different strategy in which we explicitly model the non-linear term of the dynamics with a neural network. In particular, our neural network resembles a Resnet ([He et al., 2016](#)) with the difference that, instead of the identity, the bypass is the DMD solution. Initializing the non-linear path with small random weights makes the untrained network predict the DMD solution. After training, however, results improve compared with this baseline (similar to ([Rodrigues et al., 2018](#))).

In the next sections, we present our proposed approach, which we call the Residual DMD Neural Network (ResDMD), and apply it to predict sea surface temperature (SST), comparing it to the standard DMD and a state-of-the-art dynamical model, CFSv2.

## 2. Background and Contributions

In this section, we describe the architecture of our network and the rationale for its structure. We define what the initialization is and the benefit of this choice. Finally, we describe the training procedure, which, in the current stage of this work, could be further optimized, but we leave this as future work.

<sup>1</sup>IBM Research <sup>2</sup>IBM Global Business Services. Correspondence to: Eduardo Rodrigues <edrodr@br.ibm.com>.

A multidimensional environmental variable  $\mathbf{x}$  (typically an anomaly with respect to the long-term average, i.e. the climate mean), whose components extend over the spatial domain, can be modeled as:

$$\frac{d\mathbf{x}}{dt} = L\mathbf{x} + N(\mathbf{x}) \quad (3)$$

where  $L$  is the linear part of the dynamics, and  $N$  represents nonlinear terms.  $N$  is usually, however, approximated by a linear operator, as in (Winkler et al., 2001). This approach has the major benefit of easiness to compute, since it relies on the computationally efficient singular value decomposition (SVD) algorithm. Nonetheless, one can only fully understand the system statistics by knowing the form of  $N$ .

Here, we seek to approximate the nonlinear term  $N$  as a neural network trained with stochastic gradient descent (SGD). We leverage, however, the structure and procedure for training the linear part from previous work (such as (Winkler et al., 2001) and (Newman, 2013)). The procedure for training the linear part is the Dynamic Mode Decomposition (DMD), also known as Linear Inverse Model in the atmospheric sciences.

One of the greatest advantages of the DMD method is to project the dynamics of the system being studied into a low-dimensional space, spanned by a set of spatially correlated modes that have the same linear behavior in time (Kutz et al., 2016). In discrete form the solution is given by:

$$\mathbf{x}(t_0 + n * \Delta t) = \Phi \Lambda^n \Phi^\dagger \mathbf{x}(t_0) \quad (4)$$

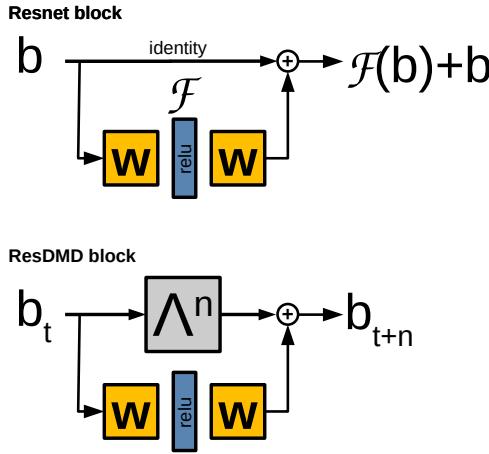


Figure 1. Comparison between Resnet block and ResDMD block.

in which  $\Phi$  and  $\Lambda$  are the eigenvectors and eigenvalues of the matrix ( $A$ ) approximating the system in Equation 3. The pseudo inverse of  $\Phi$ , i.e.  $\Phi^\dagger$ , projects the state variable  $\mathbf{x}$  into a low dimension space. Its dimension size is

given by the leading eigenvalues of  $A$ , whose number is a hyperparameter of the method.

Our proposed model also operates initially in the same low dimensional space as the DMD model. Its first step is to use the matrix  $\Phi$  to project the initial state vector, which in our case extends over the whole globe:

$$\mathbf{b} = \Phi^\dagger \mathbf{x} \quad (5)$$

This approach has the advantage of reducing the risk of overfitting with high dimensional data as the dynamics are restricted to this low dimensional sub-space. The projection matrix is an optimizable parameter in our model which can be further optimized to best fit observations during the SGD procedure.

In order to approximate the non-linear terms of the dynamics, we use a neural network architecture (with parameters  $w$ ) similar to the Resnet (He et al., 2016). The key assumption of Resnet is that it is easier to optimize the residual mapping than to optimize the original unreferenced signal. The reference in that neural network is the signal itself, i.e. an identity mapping (Figure 1 top). Our ResDMD architecture is similar, but the reference in our case is the DMD solution (Figure 1 bottom).

Initializing the weights of the projection matrices and  $\Lambda$  with the DMD matrices, and initializing the weights  $w$  to small random values effectively makes the SGD procedure look for a solution with a non-linear additive term close to the DMD solution. As one accumulates more and more observation data and trains the model, the network better approximates equation 3. In the next section, we evaluate ResDMD with SST forecasts compared to the standard DMD procedure and the CFSv2 model.

Our model can be further stacked in larger networks, similarly to the very deep Resnet models. Additional restrictions, such as shared parameters between layers for instance, can be added. In this paper, we only show results with a single block to test the viability of the method.

The training procedure is basically divided in two stages. First, we fit a DMD model to the observed data (training set). The resulting matrices  $\Phi$  and  $\Lambda$  are then used to initialize the corresponding weights of the neural network. In addition, the weights  $w$  are initialized with small random values so that the untrained network produces the DMD solution. The second stage corresponds to the SGD optimization procedure, in which we run through the training set again. In this stage, all parameters ( $\Phi$ ,  $\Lambda$  and  $w$ ) are updated.

### 3. Experiments

In order to test the performance of the ResDMD, we trained and tested it with sea surface temperature (SST). This variable is known to be an important predictor for regional climate trends and consequential weather events (Kaper & Engler, 2013), and it is reasonably slow varying making it a good candidate for a first target variable. We hereby compare performance of the ResDMD with the standard DMD and CFSv2 seasonal forecast ensemble.

The SST dataset used in our experiments is the Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5) (Huang et al., 2017). It provides estimates of SSTs from 1854 to present and is most reliable from 1950. Its spatial resolution is two degrees and is available monthly. We also use the CFSv2 seasonal forecasts as a benchmark for our predictions; it was not used as input to our model.

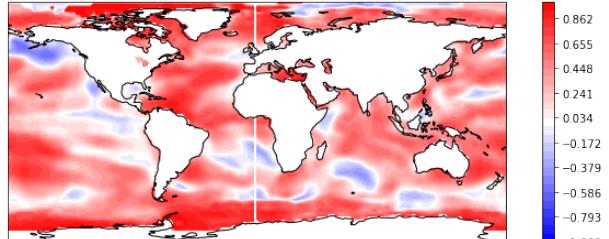
Instead of directly predicting absolute values, a common practice in long range forecasts of SST and other climate variables is to predict anomalies from the expected value (Tippett et al., 2018). (The strong seasonality in this data can bias the evaluation metric.) We computed anomalies by subtracting a monthly climate mean (from the period 1980 to 2010) from the corresponding training and testing sets. For ERSSTv5, the training set ranged from 1850-2010 and the test set from 2011 to 2020. In addition, we computed CFSv2 forecast anomalies using the CFSv2 forecast climatologies from 1980 to 2010. This extra precaution aims to make the comparison with CFSv2 more realistic – computing forecast anomalies from observed climate means (in this case ERSSTv5) could hurt the CFSv2 performance by exposing model biases (Kumar et al., 2012) and confer an unfair advantage to our model.

After training the standard DMD model and ResDMD with the training set, we evaluated the performance in the test set using the anomaly correlation coefficient (ACC), which is commonly used to measure the skill of a forecast system (Owens & Hewson, 2018). ACC values closer to unity indicate higher skill of the forecast system. We also used an ACC difference ( $\Delta$ ACC) to compare both the standard DMD and ResDMD to CFSv2. Positive values of  $\Delta$ ACC indicate the corresponding DMD model is better than CFSv2, and negative values indicate the opposite.

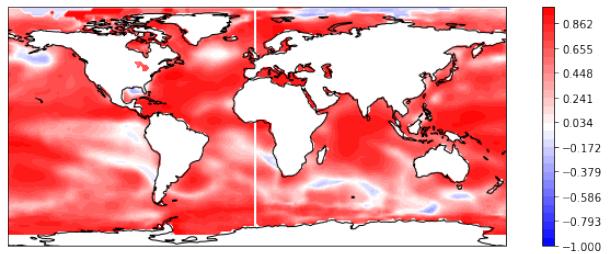
### 4. Results

In this section, we present some of the ACC maps we obtained in order to compare ResDMD with the standard DMD and CFSv2. The number of points for each spatial coordinate in those maps corresponds to the number of years in our test set (10 years) times twelve months.

Figure 2 shows ACC values for the standard DMD (top)



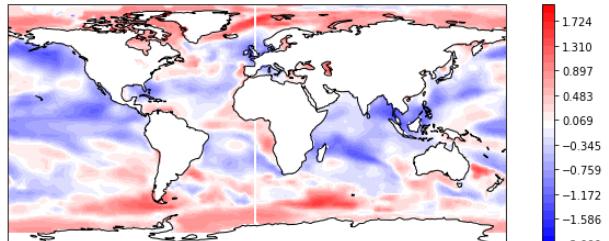
(a) Standard DMD



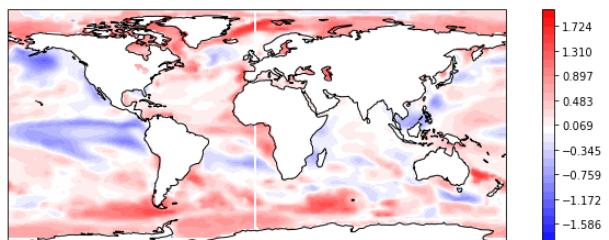
(b) ResDMD

Figure 2. ACC of SST forecast of 6 months lead time over the 2010 to 2020 period

and ResDMD (bottom) models. This particular example corresponds to predictions 6 months in advance (6-month lead time). Overall, ResDMD shows higher skill than the standard DMD. In particular regions of low correlation, such as the southern Atlantic Ocean, southern Indian Ocean and eastern Pacific Ocean, ResDMD improves skill considerably



(a) Standard DMD



(b) ResDMD

Figure 3.  $\Delta$ ACC of SST forecast of 9 months lead time over the 2011 to 2020 period compared to CFSv2

compared to the standard model. However, skill in the eastern tropical Pacific (where ENSO dynamics are observed) does not improve as much. In order to better understand this behaviour we compare results to CFSv2.

Figure 3a–3b presents the  $\Delta\text{ACC}$  for the standard DMD ACC and CFSv2, and ResDMD and CFSv2, respectively. Positive values indicate the DMD models are more skillful than the CFSv2. Here, we are only showing  $\Delta\text{ACC}$  for 9-month lead times. Again, overall one can see that ResDMD performs significantly better than the standard model. In addition, ResDMD also performs better in many places than the CFSv2. However, ResDMD has less skill in the eastern Pacific/ENSO region compared to CFSv2, a pattern that is also observed at other lead times (not shown). A possible explanation for the lack of skill in this region is that the dynamics of the state variable – explicitly simulated by CFSv2 – may be driven by a forcing term, which is strong enough to change the internal dynamics of this variable. In order to tackle this issue, we are exploring additional predictors such as 150-mb geopotential height.

Finally, we present longer range forecasts by the standard DMD and ResDMD (Figure 4) For 1-year lead time, the proposed ResDMD clearly performs better than the standard DMD across most of the globe. For 5-year lead time, the ResDMD error increases in specific locations such as the southern Indian Ocean, however it remains more skillful overall than the standard DMD. The ResDMD errors continue to amplify with lead time, and we conjecture that at longer lead times (e.g., 7-10 years) the ResDMD is no longer appropriate. We are actively exploring deeper neural networks with additional regularization for these longer lead times which will have the effect of damping the forecast close to the climate mean, however.

## 5. Final remarks

In this paper, we proposed a novel extension to the DMD method (ResDMD) that models non-linear additive terms of a dynamical system as a neural network. Weight initialization is such that, before the SGD training, the prediction is already a sensible forecast. We have shown predictions of global SSTs compared to the standard DMD and CFSv2, with notable improvements in some regions.

The objective of the ResDMD model is to make decadal forecasts which are becoming increasingly important for climate change policy and to underpin other decision-making. We believe that data-driven methods, such the one proposed here, will be an important tool along with physically-driven methods to perform such forecasts. We expect to combine these two types of methods (i.e., a hybrid physical-statistical approach) in forthcoming work to address the shortcomings of the ResDMD approach.

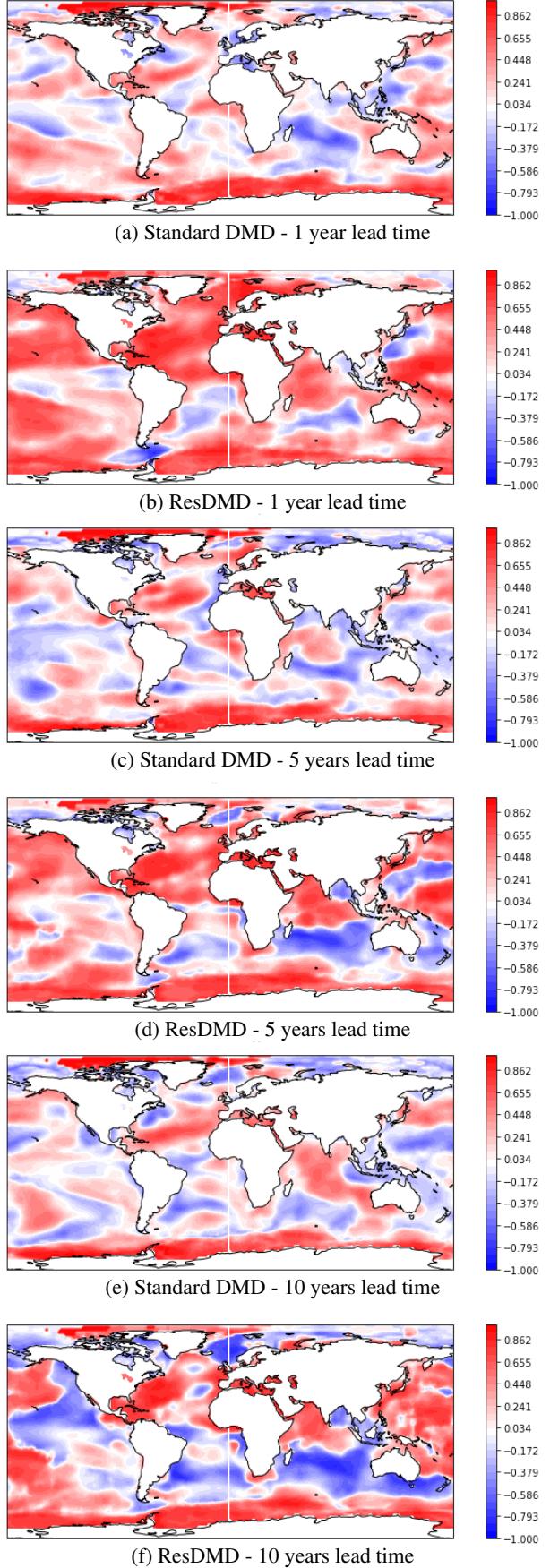


Figure 4. ACC of SST forecast

## References

- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M. Extended reconstructed sea surface temperature, version 5 (ersstv5): upgrades, validations, and intercomparisons. *Journal of Climate*, 30(20):8179–8205, 2017.
- Huddart, B. et al. Seasonal and decadal forecasts of atlantic sea surface temperatures using a linear inverse model. *Clim Dyn*, 49:1833–1845, 2017.
- Kaper, H. and Engler, H. *Mathematics and climate*. SIAM, 2013.
- Kirtman, B. P., Min, D., Infant, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M. P., Becker, E., et al. The north american multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4):585–601, 2014.
- Kumar, A., Chen, M., Zhang, L., Wang, W., Xue, Y., Wen, C., Marx, L., and Huang, B. An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the ncep climate forecast system (cfs) version 2. *Monthly weather review*, 140(9):3003–3016, 2012.
- Kutz, J. N., Brunton, S. L., Brunton, B. W., and Proctor, J. L. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.
- Newman, M. An empirical benchmark for decadal forecasts of global surface temperature anomalies. *Journal of Climate*, 26(14):5260–5269, 2013.
- Owens, R. and Hewson, T. Ecmwf forecast user guide. *Reading: ECMWF*, 10:m1cs7h, 2018.
- Rodrigues, E., Oliveira, I., Cunha, R., and Netto, M. Deep-downscale: A deep learning strategy for high-resolution weather forecast. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, pp. 415–422, 2018. doi: 10.1109/eScience.2018.00130.
- Shin, J., Park, S., Shin, S.-I., Newman, M., and Alexander, M. A. Enhancing enso prediction skill by combining model-analog and linear inverse models (ma-lim). *Geophysical Research Letters*, 47(1):e2019GL085914, 2020.
- Tippett, M. K., Trenary, L., DelSole, T., Pegion, K., and LHeureux, M. L. Sources of bias in the monthly cfsv2 forecast climatology. *Journal of Applied Meteorology and Climatology*, 57(5):1111–1122, 2018.
- Vecchi, G. Gfdl flor: Experimental high-resolution coupled seasonal to decadal prediction system, 2015. URL [https://www.gfdl.noaa.gov/wp-content/uploads/files/administrative/2015\\_marine\\_workshop/vecchi\\_06032015.pdf](https://www.gfdl.noaa.gov/wp-content/uploads/files/administrative/2015_marine_workshop/vecchi_06032015.pdf).
- Winkler, C. R., Newman, M., and Sardeshmukh, P. D. A linear model of wintertime low-frequency variability. part i: Formulation and forecast skill. *Journal of climate*, 14(24):4474–4494, 2001.
- Yeager, S., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., Karspeck, A., Lindsay, K., Long, M., Teng, H., et al. Predicting near-term changes in the earth system: A large ensemble of initialized decadal prediction simulations using the community earth system model. *Bulletin of the American Meteorological Society*, 99(9):1867–1886, 2018.