Can we predict southeastern South American precipitation anomalies using convolutional variational autoencoders?

Arianna Varuolo-Clarke | amv2185

Machine Learning for Environmental Engineering and Science
EAEE E4000 | Fall 2021

Table of Contents

Introduction & Motivation

Southeastern South America (SESA; southern Brazil, Paraguay, Uruguay, and northern Argentina) has experienced a 27% increase in precipitation from 1902-2020 during austral summer [December-February (DJF)]. This is one of the largest observed precipitation trends across the globe (Varuolo-Clarke et al., 2021). There have been increases in precipitation over the other seasons but I focus on DJF as it is SESA's rainiest season and the season with the largest and most statistically significant observed regional precipitation trend (e.g. Gonzalez et al., 2014; Varuolo-Clarke et al., 2021). As a result, the agricultural landscape has shifted in the region. There has been a 210% expansion in the area of Argentina's soy cultivation over the last several decades (i.e., Baldi & Paruelo, 2008; Barreiro et al., 2014; Lucas et al., 2018; Magrin et al., 2005). Additionally, Brazil, Argentina, and Paraguay make up the 2nd, 3rd, and 6th largest soy producers globally (http://www.fao.org/faostat/en/#rankings/commodities_by_country). While the increase in precipitation over SESA has been well documented (e.g., Carvalho, 2020; Dai, 2021; de Barros Soares et al., 2017; Haylock et al., 2006; Liebmann et al., 2004; Zilli et al., 2017), there is still uncertainty in understanding what is driving the trend over the historical interval from 1902-2020. Furthermore, climate models appear unable to simulate similar precipitation trends in the region in fully-coupled historical simulations and sea surface temperature (SST) forced simulations from Coupled Model Intercomparison Project Phase 3 (CMIP3), CMIP5, and CMIP6 (i.e., Díaz et al., 2021; Gonzalez et al., 2014; Seager et al., 2010; Varuolo-Clarke et al., 2021; Zhang et al., 2016). However, climate models do project precipitation will increase in SESA due to greenhouse gas emissions (e.g., Cook et al., 2020; Varuolo-Clarke et al., 2021), making it critical to better understand the key drivers at play.

Seager et al. (2010) has shown that wet conditions in SESA are likely to occur under positive El Niño Southern-Oscillation (ENSO) conditions and the cool phase of Atlantic multidecadal variability (AMV). Gonzalez et al. (2014) found the increase in SESA precipitation associated with ozone depletion was as large, or larger, than the trend associated with GHG forcing post-1960 in various National Center for Atmospheric Research simulations. However, using the Geophysical Fluid Dynamics Lab model, Zhang et al. (2016), found precipitation increased over SESA due to GHG emissions and not from aerosols or stratospheric ozone depletion. Given the apparent lack of model ability to simulate the observed increase in precipitation over SESA and the competing forcings that play a role in setting the hydroclimate in this region, *I am motivated*

to use machine learning techniques to identify better predictors of SESA precipitation? Specifically, I will assess the ability of neural networks to predict SESA precipitation anomalies given several different climate indices.

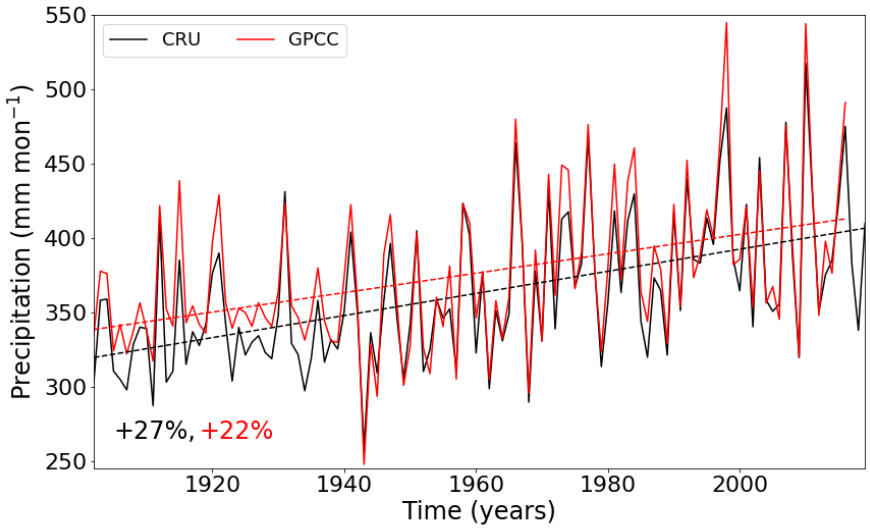

Figure 1. SESA precipitation from CRU (1902-2020) and GPCC (1902-2018).

Data

My goal is to understand drivers of precipitation over SESA using a variety of different climate indicators. All data used are DJF monthly means from 1949-2020. For precipitation , I use version 4.05 of the Climatic Research Unit dataset (CRU TSv4.05; Harris et al., 2020) that is on a 0.5° land-only grid. I seek to predict the SESA precipitation anomaly relative the 1949-2020 mean. The climate indices based on sea surface temperatures (SSTs) are from the NOAA NCDC Extended Reconstructed SST product (ERSSTv5; Huang et al., 2020) on a 2.0° grid. Using ERSSTv5, I get SST data from the NINO3.4 region for DJF as a metric the El Niño Southern Oscillation (ENSO, 5°S-5°N, 120°W-170°). I also get SSTs over the North Pacific as a metric of the Pacific Decadal Oscillation (PDO, 20°N-90°N, 120°W-160°E). In the Atlantic, I get SST data over the North Atlantic for a metric of the Atlantic Multidecadal Variability (AMV, 0°-65°N, 80°W-65°E), for a region over the southern North Atlantic (SNA, 0°-10°N, 20°W-50°W) and lastly a gradient of SST defined as the difference between SSTs at 45°S, 20°W and 20°S, 20°W. Atmospheric variables are reanalysis data from the National Centers for Environmental Prediction and Atmospheric Research (NNR; Kalnay et al., 1996). I use 500mb geopotential height from NNR to calculate an index for the Southern Annular Mode (SAM) for DJF. Finally, indices of the

South American low-level jet that I have developed are used for metrics of total moisture flux (TMF), wind-driven moisture flux (WDMF), humidity-driven moisture flux (HDMF), and the residual term (Q*V interaction) jet indices are also evaluated as contributors to SESA precipitation anomalies (Varuolo-Clarke et al., *under review*).

Dataset	Full name	Data source
1. ENSO	El Niño Southern oscillation	ERSSTv5
2. PDO	Pacific decadal oscillation	ERSSTv5
3. AMV	Atlantic multidecadal variability	ERSSTv5
4. SNA	Southern north Atlantic	ERSSTv5
5. GRAD	Southern Atlantic gradient	ERSSTv5
6. TMF	Total moisture flux	NNR
7. WDMF	Wind-driven moisture flux	NNR
8. HDMF	Humidity-driven moisture flux	NNR
9. QV	Q*V interaction	NNR
10. SAM	Southern annular mode	NNR
11. PR*	Precipitation	CRU

Table 1. Information on inputs and the precipitation output* for the neural network.

Machine Learning Methods

Neural networks

For this problem of predicting precipitation anomalies, a continuous value, given data about oceanic and atmospheric data points around the globe, I chose to use a neural network, more specifically a sequential model from the TensorFlow package.

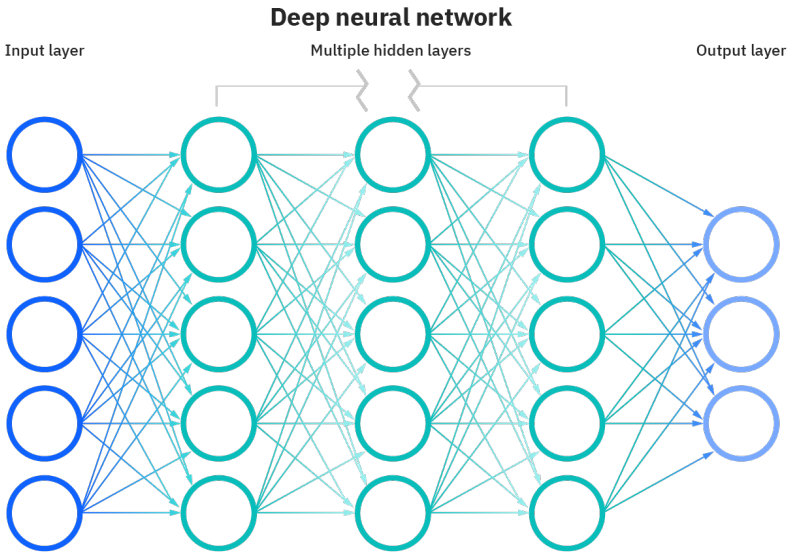

Figure 2. Sketch of neural networks with inputs, hidden layers, and the output. Figure taken from IBM.com

In order to process and prepare the data accordingly to follow the Boston Housing Prices class example, I set my data up so that it is in a Pandas Dataframe (Figure 3). There are 72 years of data and 11 columns in total, with the last column (PR) being our predicted field, precipitation anomalies. I check that there are no missing values in the data and since there are none, I move on to looking at the data (Figure 4).

```
[12]: predictors['PR'] = pr['PR']
      predictors
```

[12]:

	ENSO	PDO	AMV	SNA	GRAD	TMF	WDMF	HDMF	QV	SAM	PR
0	25.889908	-1.682120	0.055977	25.889908	-0.393548	-63.756079	-61.436730	-6.346638	4.027290	-2.264632	-77.771730
1	24.644358	-1.656283	-0.002819	24.644358	-1.160739	-1.661660	-1.997012	-0.651827	0.987180	-1.846696	-52.998200
2	25.366077	-1.739386	0.071659	25.366077	1.964946	5.257056	6.081561	-0.257847	-0.566658	-0.699164	17.117737
3	26.729053	-1.698638	0.058565	26.729053	-1.532392	-5.095347	-4.735496	1.039970	-1.399819	-1.751757	-63.901855
4	26.601658	-1.910678	0.222424	26.601658	-0.217099	0.897549	3.373979	0.335203	-2.811634	-1.705375	-67.280270
...	...	...	...	...	...	...	...	...	...	...	...
67	29.093943	-1.867435	0.232822	29.093943	-1.374376	37.100743	25.842593	6.843624	4.414527	0.788260	125.100525
68	26.244295	-1.882998	0.305046	26.244295	-0.799298	19.503688	12.921880	3.953870	2.627938	-0.003212	15.529511
69	25.634764	-1.618691	0.261969	25.634764	2.004310	5.351652	-0.297280	5.124919	0.524013	0.977473	-38.272583
70	27.314991	-1.733327	0.069532	27.314991	-0.252742	16.627084	10.840434	3.317132	2.469518	0.700910	30.939514
71	27.053848	-1.780369	0.317152	27.053848	0.499439	3.348179	-3.863303	5.368122	1.843360	-0.116220	-20.987396

72 rows × 11 columns

Figure 3. Screen shot of Dataframe showing the inputs and label/output data (precipitation, PR).

```
[33]: predictors.isnull().sum()

[33]: ENSO    0
      PDO     0
      AMV     0
      SNA     0
      GRAD    0
      TMF     0
      WDMF    0
      HDMF    0
      QV      0
      SAM     0
      PR      0
      dtype: int64
```

Figure 4. I check if there are any null values in my Dataframe and find that there are no missing values.

Figure 5 is the distribution of the target precipitation anomaly variable. The values are not quite distributed normally and this makes sense given that I am actually trying to assess the increase in precipitation in this region. However, I had not looked at the precipitation anomalies like this. It is neat to see the somewhat bimodal distribution of anomalies that are negative vs. anomalies that are positive. Other things to note are that the positive trends are just about below 150 mm mon^{-1}, while the negative anomalies are never more than -100 mm mon^{-1}.

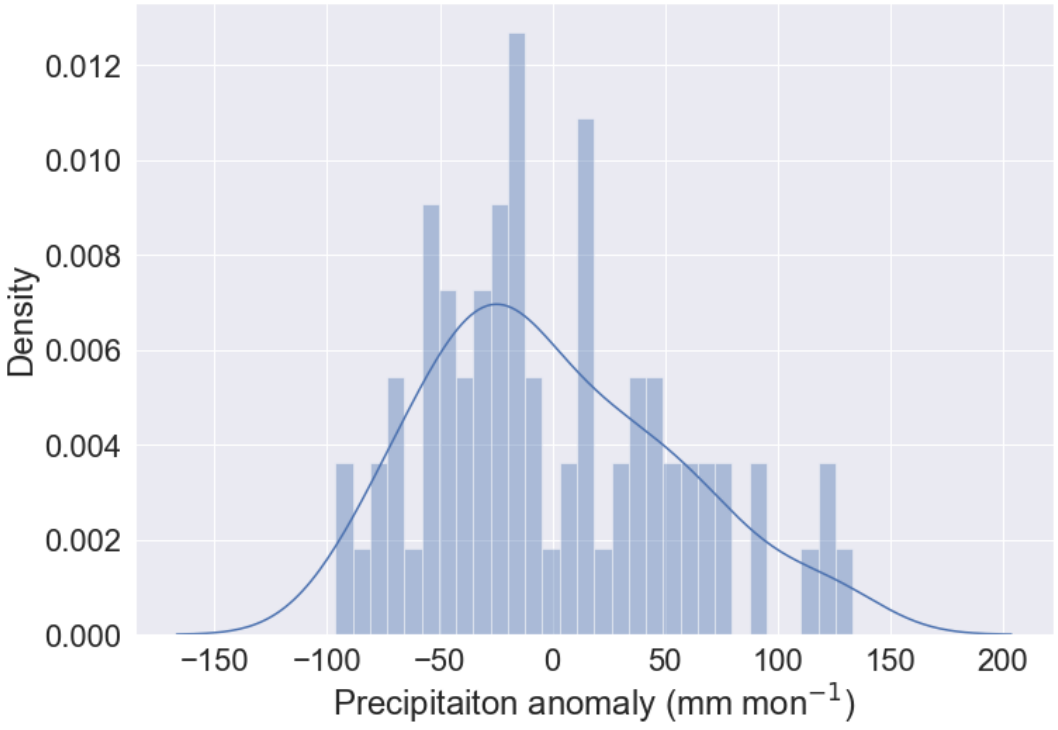

Figure 5. Distribution of DJF SESA precipitation anomalies from 1949-2020.

Figure 6 presents the linear relationship between all of the variables as a correlation matrix. The TMF and WDMF are highly correlated with one another (r = 0.98) so I drop WDMF and just use TMF. I also notice that the AMV and southern North Atlantic SSTs are also rather high (r = 0.78). This is due to the overlap in the region of the SSTs and I chose to keep both of these as inputs anyway.

There is also a strong correlation between HDMF and the southern North Atlantic as this value is over the South American low-level jet region which receives a lot of moisture from the Atlantic. Therefore, it makes sense that there would be a strong correlation between these two variables. For the purpose of this project, I decided to keep both of these as inputs. The precipitation data is most highly correlated with TMF (r = 0.65), WDMF (r = 0.62), ENSO (r = 0.47), and HDMF (r = 0.42).

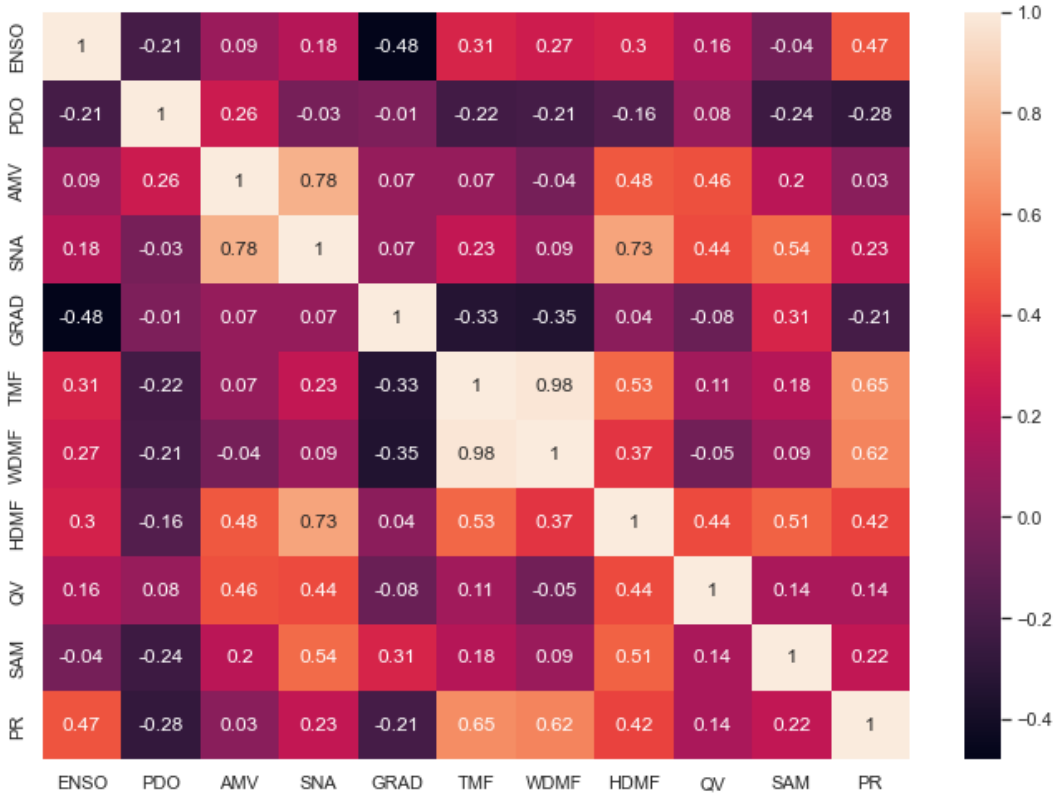

Figure 6. Correlation matrix of input and predictor for the neural network.

Preparing the Neural Network

For this project, I will assess the ability of 10 different inputs to predict SESA precipitation anomalies using a neural network. I will start with linear activation functions before turning to the Relu activation function, with the hope that the Relu activation function will actually improve over the linear activation function as using the linear activation function is essentially just a linear regression.

Build the model

```
[45]: model = keras.Sequential([
          keras.layers.Dense(64, activation=tf.keras.activations.linear,
                             input_shape=(train_data.shape[1],),
                             name='hidden_layer_1',),
          keras.layers.Dense(64, activation=tf.keras.activations.linear, #activation=tf.nn.relu,
                             name='hidden_layer_2',),
          keras.layers.Dense(1, name='output_layer',)
      ])

      # Define optimizer
      optimizer = tf.optimizers.Adam()

      model.compile(loss='mse',
                    optimizer=optimizer,
                    metrics=['mae'])
```

Figure 7. Building the model. I start with two layers, with 64 neurons for each layer and the linear activation function.

```
[46]: model.summary()
```

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
hidden_layer_1 (Dense)       (None, 64)                640
_________________________________________________________________
hidden_layer_2 (Dense)       (None, 64)                4160
_________________________________________________________________
output_layer (Dense)         (None, 1)                 65
=================================================================
Total params: 4,865
Trainable params: 4,865
Non-trainable params: 0
_________________________________________________________________
```

Figure 8. Compile the model. Here is a summary of the two layer with 64 neurons model showing the types of layers, their output, shape, and number of parameters. In total there are 4,865 trainable parameters.

After exploring the neural networks with 64 neurons, I move down to 16 neurons for each layer. Next, I move to using the Relu activation function. I start out with two layers and either 16, 64, or 128 neurons and then I increase the depth of layers from two to three with 16 or 64 neurons.

Finally, I explore what happens with four or five layers and 16 neurons. In all cases shown here I use 1000 epochs, and a patience of 20 for the early stopping function. For each of the results shown below I consider five different runs given the same model parameters.

Results & Discussion

Linear regression

Starting with 64 neurons and two layers over five simulations, I get an average mean absolute error (MAE) of 36.27 and a correlation coefficient of 0.48 (Table 2). However, when looking at a neural network with 16 neurons instead of 64, all else remaining equal, the average MAE for five simulations is 36.1 and the correlation coefficient increases to 0.5 (Table 3). Perhaps a reason for some improvements is when considering two layers of 64 neurons, there are nearly 5,000 trainable parameters. However, when we consider two layers of 16 neurons, there are only 449 trainable parameters which is much closer to the order of magnitude of the inputs I am giving the model (10 inputs for 72 years is ~700 data points). Given that two layers 16 neurons have more similar order of magnitude to the inputs, this makes sense that we might be getting a better model when considering only 16 neurons as opposed to 64. What happens when we leave the linear space and introduce the Relu activation function?

2 linear layers, 1 outer layer, **64 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	36.35	0.48	-0.13
2.	37.21	0.49	-0.17
3.	34.59	0.44	-0.11
4.	37.52	0.50	-0.17
5.	35.66	0.48	-0.11
Average	**36.27**	**0.48**	**-0.14**

Table 2. Metrics from 5 simulations with 2 linear layers of a size 64 neuron neural network.

2 linear layers, 1 outer layer, **16 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	35.29	0.50	-0.04
2.	36.15	0.50	-0.05
3.	36.42	0.49	-0.07
4.	36.22	0.50	-0.02
5.	36.26	0.49	-0.07
Average	**36.1**	**0.50**	**-0.05**

Table 3. Metrics from 5 simulations with 2 linear layers of a size 64 neuron neural network.

Introducing non-linearity

Next, I consider two layers, each with 64 neurons, and the Relu activation function. I find the average MAE is 36.07 mm mon^{-1} with a correlation coefficient of 0.51 (Table 4). This is an improvement over using the linear activation function under otherwise identical configurations. When I decrease the neuron size from 64 to 16 or increase it from 64 to 128 the results are worse (Tables 5 and 6). I am not sure what to make of these results and continue to explore this architecture by increasing the depth of the neural network.

2 relu layers, 1 outer layer, **64 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	36.88	0.50	-0.21
2.	34.53	0.52	-0.04
3.	35.82	0.48	-0.20
4.	35.17	0.49	-0.16
5.	39.31	0.44	-0.45
Average	**36.34**	**0.49**	**-0.21**

Table 4. Metrics from 5 simulations with 2 Relu layers of a size 64 neuron neural network.

2 relu layers, 1 outer layer, **16 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	35.16	0.41	-0.17
2.	41.97	0.47	-0.39
3.	32.20	0.56	0.13
4.	39.47	0.50	-0.23
5.	37.77	0.50	-0.23
Average	**37.3**	**0.49**	**-0.178**

Table 5. Metrics from 5 simulations with 2 Relu layers of a size 16 neuron neural network.

2 relu layer, 1 outer layer, **128 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	36.60	0.50	-0.2
2.	41.91	0.50	-0.43
3.	34.27	0.53	-0.01
4.	37.86	0.50	-0.18
5.	35.73	0.52	-0.11
Average	**37.27**	**0.51**	**-0.19**

Table 6. Metrics from 5 simulations with 2 Relu layers of a size 128 neuron neural network.

Increasing depth of the neural network

Next, considering the Relu activation function, I increase the depth of my neural network from two layers to three and consider 64 neurons, 16 neurons, and 128 neurons. In the case of 64 neurons and three layers, the average MAE 40.54, which is the worst MAE error I have seen so far (Table 7). The correlation value is also quite low at 0.40. When decreasing the neurons down to 16 neurons three is some improvement overall. The average MAE for the five simulations with this configuration drops to 35.77, the lowest MAE I have seen (Table 8). Similarly, the correlation coefficient is now 0.51, still not very high but the highest observed. With 128 neurons and three layers the average MAE becomes 40.10 (Table 9) showing some improvement over 64 neurons but a degradation compared to 16 neurons. The average correlation coefficient for 128 neurons is also the lowest across this set of configurations at 0.38. The next step is to continue investigating if there is an improvement in predicting SESA precipitation anomalies considering 16 neurons but four or five layers.

3 relu layer, 1 outer layer, **64 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	35.49	0.48	-0.14
2.	37.99	0.39	-0.35
3.	43.53	0.35	-0.71
4.	41.64	0.43	-0.41
5.	44.07	0.38	-0.68
Average	**40.54**	**0.40**	**-0.46**

Table 7. Metrics from 5 simulations with 3 Relu layers of a size 64 neuron neural network.

3 relu layer, 1 outer layer, **16 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	31.78	0.52	0.09
2.	40.06	0.48	-0.28
3.	31.36	0.49	0.06
4.	37.40	0.51	-0.14
5.	38.26	0.54	-0.11
Average	**35.77**	**0.51**	**-0.08**

Table 8. Metrics from 5 simulations with 3 Relu layers of a size 16 neuron neural network.

3 relu layer, 1 outer layer, **128 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	44.91	0.32	-0.68
2.	41.94	0.44	-0.45
3.	38.58	0.46	-0.33
4.	42.14	0.41	-0.48
5.	38.57	0.46	-0.29
Average	**41.23**	**0.42**	**-0.45**

Table 9. Metrics from 5 simulations with 3 Relu layers of a size 128 neuron neural network.

Considering four Relu layers with 16 neurons each, we actually see a degradation in the average MAE which is 38.1, with the correlation coefficient falling bellowing 0.5 at 0.46. Furthermore, when we consider five layers of 16 neurons, we find the average MAE actually increases from 38.1 (four layers) to 39.27 (Table 11). Interestingly, the correlation coefficient is also 0.46 under this configuration.

4 relu layer, 1 outer layer, **16 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	39.03	0.43	-0.22
2.	42.78	0.39	-0.61
3.	39.74	0.46	-0.35
4.	36.95	0.52	-0.16
5.	31.76	0.49	0.12
Average	**38.1**	**0.46**	**-0.24**

Table 10. Metrics from 5 simulations with 4 Relu layers of a size 16 neuron neural network.

5 relu layer, 1 outer layer, **16 neurons**, 1000 epochs, 20 patience

# of simulations	Testing set MAE	Correlation coeff.	r_2 score
1.	42.65	0.41	-0.60
2.	31.68	0.55	0.16*
3.	38.87	0.53	-0.19*
4.	40.67	0.36	-0.39
5.	42.49	0.46	-0.51*
Average	**39.27**	**0.46**	**-0.30**

Table 11. Metrics from 5 simulations with 5 Relu layers of a size 16 neuron neural network.

Exploring 16 neurons and varying Relu layers

Now that I have explored linear vs. Relu activation functions and 16, 64, or 128 neurons, I would like to focus in on 16 neurons and the differences across two layers through five layers.

From two through five layers, the best average MAE value is 35.77 which occurs with three layers. This is also when the best correlation coefficient value occurs at 0.51. Figure 9 presents the training vs. validation loss (left), the scatter plot of the actual precipitation values vs. the predicted precipitation (center) and the distribution of the prediction error (right).

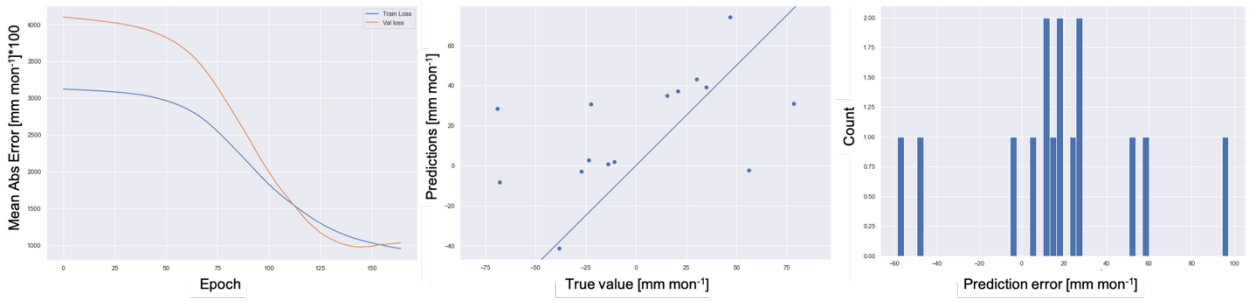

Figure 9. Model parameters. Training loss vs. validation loss (left). Predictions of precipitation vs. true values of precipitation, correlation coefficient = 0.52 (center). Distribution of prediction error (right).

I present the same figures for Figure 10-14 and some of the highlights are presented in this section. In most cases the training loss and validation loss flatten between 100 and 200 epochs except for in the 5th simulation, when the losses don't flatten until ~350 epochs. In the center panels of the predictions vs. true values we see that the correlations are never really that strong, they all are near 0.5, which is not very high.

I am very interested in the differences in the prediction error distribution. In the first scenario, the peak distribution is between 10-30 mm mon^{-1}, with a count of ~10 falling in this range. In the next scenario, there is only one peak with a count of 2 with its peak at 20 mm mon^{-1}. The second scenario is in contrast to the third where there are four peaks spanning much of the distribution. In the fourth scenario, we see the highest peak of 3 counts centered on ~25 mm mon^{-1}. Lastly, he fifth scenario has one peak at ~25 mm mon^{-1} as well. Therefore, across all of these scenarios, there is not a lot of convergence on the prediction error leading me to consider what are the ways that I could improve the predictions of SESA precipitation anomalies?

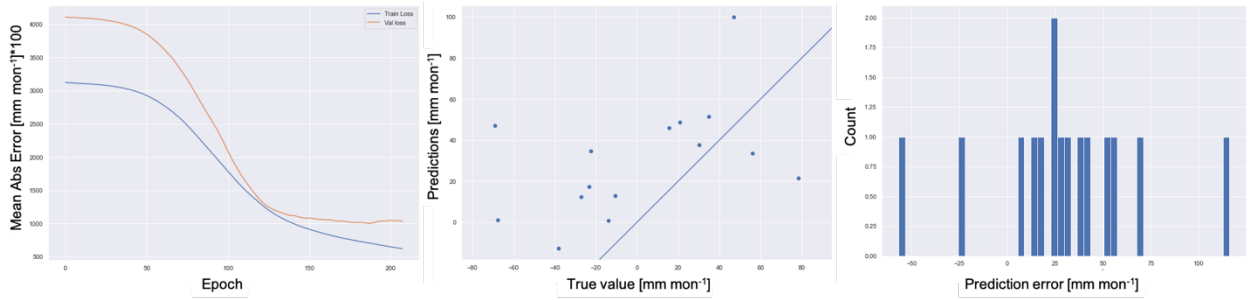

Figure 10. Model parameters. Training loss vs. validation loss (left). Predictions of precipitation vs. true values of precipitation, correlation coefficient = 0.48 (center). Distribution of prediction error (right).

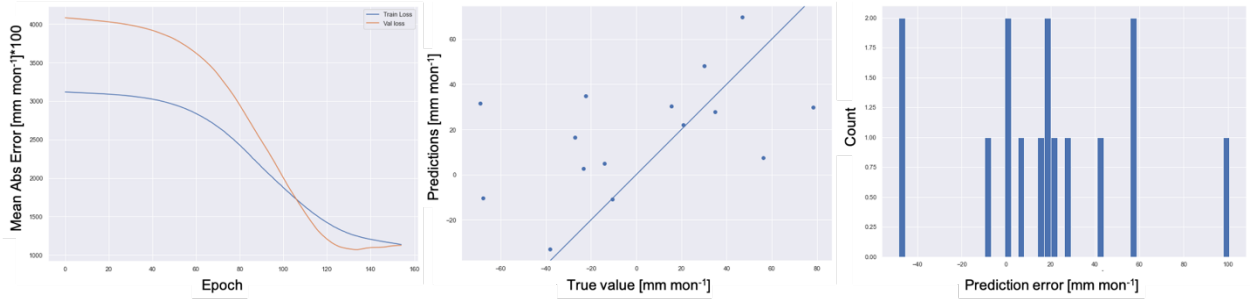

Figure 11. Model parameters. Training loss vs. validation loss (left). Predictions of precipitation vs. true values of precipitation, correlation coefficient = 0.49 (center). Distribution of prediction error (right).

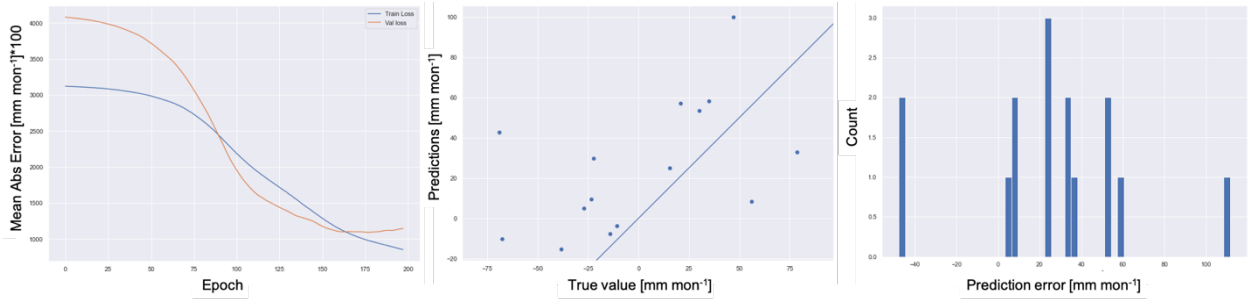

Figure 12. Model parameters. Training loss vs. validation loss (left). Predictions of precipitation vs. true values of precipitation, correlation coefficient = 0.51 (center). Distribution of prediction error (right).

Figure 13. Model parameters. Training loss vs. validation loss (left). Predictions of precipitation vs. true values of precipitation, correlation coefficient = 0.54 (center). Distribution of prediction error (right).

Summary & future work

For this project, I tested the ability of 10 different climate indicators to predict SESA precipitation anomalies. I used SSTs from different regions over the Pacific and Atlantic Oceans as well as atmospheric variables over South America and for the Southern Annular mode. I consider 72 years of December through February monthly averaged means from 1949-2020. The code for this project can be found on GitHub at https://github.com/amvaruolo-clarke/machine-learning-project.

I use convolutional neural networks, starting with a linear activation function and 16 neurons or 64 neurons and two layers. Comparing these two configurations I find that 16 neurons offer some improvement over 64 neurons. Moving on to the Relu activation function instead of a linear function, I compare 16 neurons to 64 or 128 neurons all with two layers each. In opposition to the results with the linear activation function, considering the Relu activation function, 64 neurons outperform 16 and 64 neurons. Next, I increase the number of layers to three and for the three different neuron sizes find that a size of 16 neurons yields the best results. Then, considering 16 neurons, I test the architecture with increasing depth and run the model with four Relu layers and five Relu layers. This does not necessarily yield much improvement and in fact, 16 neurons and three Relu layers yield the best results with an average MAE of 35.77 and a correlation coefficient of 0.51.

Through these experiments given this specific architecture, I have unfortunately not been able find a correlation coefficient between the predictors and the true value larger than 0.51. There are several different ways to interpret this and different ways to improve the model. First off, I did not change the learning rate of the optimizer from the default and I do wonder what kind of

improvement or degradation that would have had for the prediction of SESA precipitation anomalies. Another idea that would have potentially fruitful would have been to look at neurons smaller than 16. I do feel that 16 is likely the best option since then the trainable parameters are roughly equal to the inputs. But still, further testing out different neuron sizes might yield better results.

There are also other techniques that I could employ within this framework of time series analysis. Namely, causal discovery, which is a method to distinguish between direct and indirect dependencies and common drivers across many time series going beyond just inferring directionality between two time series (Runge, 2018).

It is also possible that I missed indicators of SESA precipitation anomalies and this may be part of the reason the correlation coefficients are not that high. If I were to repeat this analysis, I would like to include a metric for the Indian Ocean to assess its influence on SESA precipitation anomalies. Another key piece is that I only assessed 72 years of data, but what I am truly after is a ~120-year trend. Due to data constraints for some of the variables assessed here, I constrained everything to the 72-year period. A next step then would be to complete this analysis again over the longer interval excluding the variables only available over the shorter time period.

Finally, I am really excited by the abilities of a conditional variational autoencoder (CVAE) and what that will allow me to do to take this research to the next level and I plan to assess this technique further to better understand it so that I can employ it in my research. Ideally, I would like to use CVAE to see if I can use geopotential height (GPH) anomalies to reconstruct SESA precipitation anomalies. I would assess the ability of global GPH anomalies, hemisphere-only, or just over South America to reconstruct SESA precipitation anomalies.

References

Carvalho, L. M. V. (2020). Assessing precipitation trends in the Americas with historical data: A review. *WIREs Climate Change, 11*(2), e627. https://doi.org/https://doi.org/10.1002/wcc.627

Cook, B. I., Mankin, J. S., Marvel, K., Williams, A. P., Smerdon, J. E., & Anchukaitis, K. J. (2020, Jun 2020
2020-06-25). Twenty-First Century Drought Projections in the CMIP6 Forcing Scenarios. *Earth's Future, 8*(6). https://doi.org/http://dx.doi.org/10.1029/2019EF001461

Dai, A. (2021, 2021/02/21). Hydroclimatic trends during 1950–2018 over global land. *Climate Dynamics*. https://doi.org/10.1007/s00382-021-05684-1

de Barros Soares, D., Lee, H., Loikith, P. C., Barkhordarian, A., & Mechoso, C. R. (2017). Can significant trends be detected in surface air temperature and precipitation over South America in recent decades? *International Journal of Climatology, 37*(3), 1483-1493. https://doi.org/https://doi.org/10.1002/joc.4792

Díaz, L. B., Saurral, R. I., & Vera, C. S. (2021). Assessment of South America summer rainfall climatology and trends in a set of global climate models large ensembles. *International Journal of Climatology, 41*(S1), E59-E77. https://doi.org/https://doi.org/10.1002/joc.6643

Gonzalez, P. L. M., Polvani, L. M., Seager, R., & Correa, G. J. P. (2014, 2014/04/01). Stratospheric ozone depletion: a key driver of recent precipitation trends in South Eastern South America. *Climate Dynamics, 42*(7), 1775-1792. https://doi.org/10.1007/s00382-013-1777-x

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020, 2020/04/03). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data, 7*(1), 109. https://doi.org/10.1038/s41597-020-0453-3

Haylock, M. R., Peterson, T. C., Alves, L. M., Ambrizzi, T., Anunciação, Y. M. T., Baez, J., Barros, V. R., Berlato, M. A., Bidegain, M., Coronel, G., Corradi, V., Garcia, V. J., Grimm, A. M., Karoly, D., Marengo, J. A., Marino, M. B., Moncunill, D. F., Nechet, D., Quintana, J., Rebello, E., Rusticucci, M., Santos, J. L., Trebejo, I., & Vincent, L. A. (2006). Trends in Total and Extreme South American Rainfall in 1960–2000 and Links with Sea Surface Temperature. *Journal of Climate, 19*(8), 1490-1512. https://doi.org/10.1175/jcli3695.1

Huang, H., Xue, Y., Chilukoti, N., Liu, Y., Chen, G., & Diallo, I. (2020, 15 Oct. 2020). Assessing Global and Regional Effects of Reconstructed Land-Use and Land-Cover Change on Climate since 1950 Using a Coupled Land–Atmosphere–Ocean Model. *Journal of Climate, 33*(20), 8997-9013. https://doi.org/10.1175/jcli-d-20-0108.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., & Joseph, D. (1996, 01 Mar. 1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society, 77*(3), 437-472. https://doi.org/10.1175/1520-0477(1996)077<0437:Tnyrp>2.0.Co;2

Liebmann, B., Vera, C. S., Carvalho, L. M. V., Camilloni, I. A., Hoerling, M. P., Allured, D., Barros, V. R., Báez, J., & Bidegain, M. (2004). An Observed Trend in Central South American Precipitation. *Journal of Climate, 17*(22), 4357-4367. https://doi.org/10.1175/3205.1

Runge, J. (2018). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 28*(7), 075310.

Seager, R., Naik, N., Baethgen, W., Robertson, A., Kushnir, Y., Nakamura, J., & Jurburg, S. (2010, 2010 Oct 15
2019-02-06). Tropical Oceanic Causes of Interannual to Multidecadal Precipitation Variability in Southeast South America over the Past Century*. *Journal of Climate, 23*(20), 5517-5528,5534,5536-5539. http://ezproxy.cul.columbia.edu/login?url=https://search.proquest.com/docview/807443447?accountid=10226
http://rd8hp6du2b.search.serialssolutions.com/directLink?&atitle=Tropical+Oceanic+Causes+of+Interannual+to+Multidecadal+Precipitation+Variability+in+Southeast+South+America+over+the+Past+Century*&author=Seager%2C+Richard%3BNaik%2C+Naomi%3BBaethgen%2C+Walter%3BRobertson%2C+Andrew%3BKushnir%2C+Yochanan%3BNakamura%2C+Jennifer%3BJurburg%2C+Stephanie&issn=08948755&title=Journal+of+Climate&volume=23&issue=20&date=2010-10-15&spage=5517&id=doi:&sid=ProQ_ss&genre=article

Varuolo-Clarke, A. M., Smerdon, J. E., Williams, A. P., & Seager, R. (2021). Gross Discrepancies between Observed and Simulated Twentieth-to-Twenty-First-Century Precipitation Trends in Southeastern South America. *Journal of Climate, 34*(15), 6441-6457.

Zhang, H., Delworth, T. L., Zeng, F., Vecchi, G., Paffendorf, K., & Jia, L. (2016, Dec 2016
2019-04-29). Detection, Attribution, and Projection of Regional Rainfall Changes on (Multi-) Decadal Time Scales: A Focus on Southeastern South America. *Journal of Climate, 29*(23), 8515-8534. https://doi.org/http://dx.doi.org/10.1175/JCLI-D-16-0287.1

Zilli, M. T., Carvalho, L. M. V., Liebmann, B., & Silva Dias, M. A. (2017). A comprehensive analysis of trends in extreme precipitation over southeastern coast of Brazil. *International Journal of Climatology, 37*(5), 2269-2279. https://doi.org/https://doi.org/10.1002/joc.4840