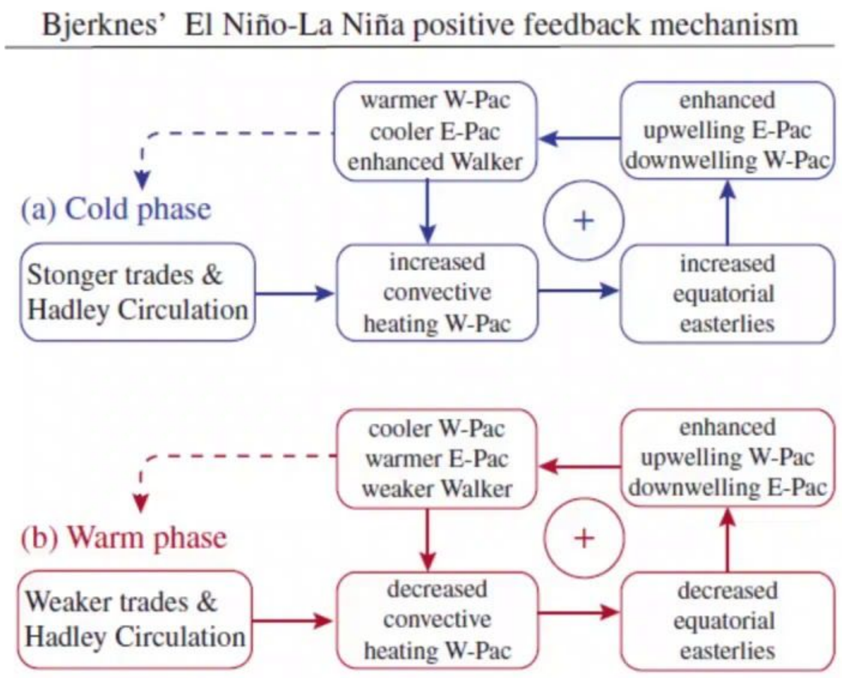
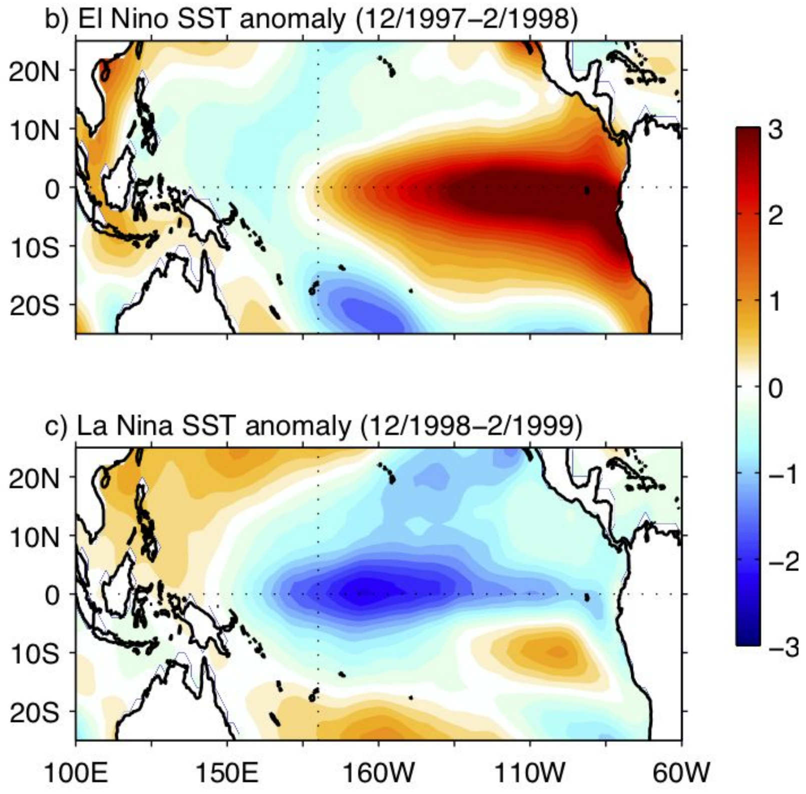
ENSO Prediction Using Deep Learning

Lilian Zhu
Anomaly Detection Aces
Milestone 1 Presentation

Project Description/Previous Solutions

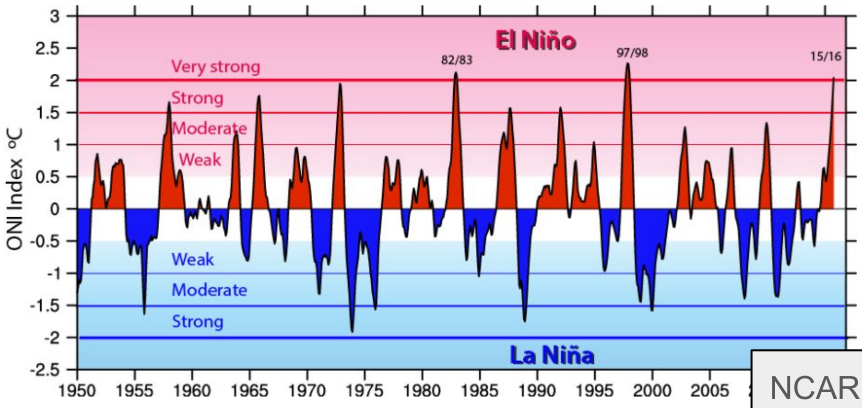
El Nino-Southern Oscillation (ENSO)



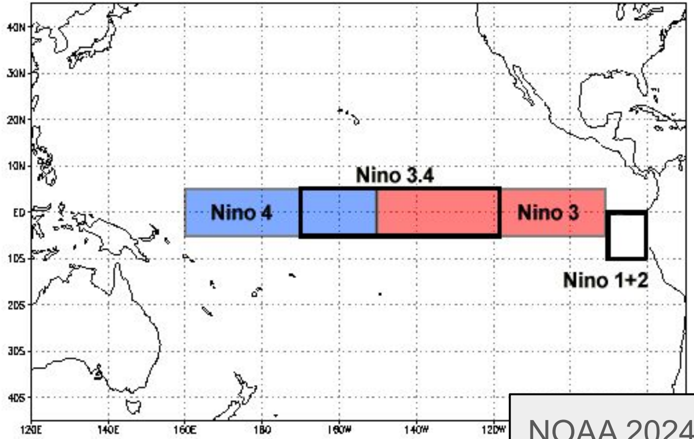


NCAR 2019

ONI - Defining Phases of ENSO



NCAR 2024



NOAA 2024

ONI (Oceanic Nino Index) = 3 month Rolling mean of SST anomaly in NINO3.4 region

- El Nino: ONI > 0.5
- La Nina: ONI < -0.5

ONI from 2012 to 2024

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2012	-0.9	-0.7	-0.6	-0.5	-0.3	0.0	0.2	0.4	0.4	0.3	0.1	-0.2
2013	-0.4	-0.4	-0.3	-0.3	-0.4	-0.4	-0.4	-0.3	-0.3	-0.2	-0.2	-0.3
2014	-0.4	-0.5	-0.3	0.0	0.2	0.2	0.0	0.1	0.2	0.5	0.6	0.7
2015	0.5	0.5	0.5	0.7	0.9	1.2	1.5	1.9	2.2	2.4	2.6	2.6
2016	2.5	2.1	1.6	0.9	0.4	-0.1	-0.4	-0.5	-0.6	-0.7	-0.7	-0.6
2017	-0.3	-0.2	0.1	0.2	0.3	0.3	0.1	-0.1	-0.4	-0.7	-0.8	-1.0
2018	-0.9	-0.9	-0.7	-0.5	-0.2	0.0	0.1	0.2	0.5	0.8	0.9	0.8
2019	0.7	0.7	0.7	0.7	0.5	0.5	0.3	0.1	0.2	0.3	0.5	0.5
2020	0.5	0.5	0.4	0.2	-0.1	-0.3	-0.4	-0.6	-0.9	-1.2	-1.3	-1.2
2021	-1.0	-0.9	-0.8	-0.7	-0.5	-0.4	-0.4	-0.5	-0.7	-0.8	-1.0	-1.0
2022	-1.0	-0.9	-1.0	-1.1	-1.0	-0.9	-0.8	-0.9	-1.0	-1.0	-0.9	-0.8
2023	-0.7	-0.4	-0.1	0.2	0.5	0.8	1.1	1.3	1.6	1.8	1.9	2.0
2024	1.8	1.5	1.1	0.7	0.4	0.2						

ENSO Prediction Errors

Historically, ENSO prediction beyond 6 months is very inaccurate.

Rapid Development of Systematic ENSO-Related Seasonal Forecast Errors

J. D. Beverley[1,2], M. Newman[2], and A. Hoell[2]

[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO, USA, [2]NOAA Physical Sciences Laboratory, Boulder, CO, USA

Abstract Climate models exhibit known systematic errors in their representation of the El Niño-Southern Oscillation (ENSO). In this study, we show that such simulation errors are largely present in tropical seasonal prediction, even for short lead times. Regressing monthly forecast errors from 11 different operational models upon the observed ENSO state, we find that predicted ENSO-related sea surface temperature anomalies (of either sign) for winter/spring are significantly extended or shifted to the west and are also too persistent during the ENSO decay phase, both common climate model errors. There are also corresponding precipitation forecast errors, most notably a robust westward shift of the ENSO-related precipitation dipole that may impact predictions of extratropical teleconnections. These ENSO-related errors develop within days after initialization regardless of month, including significant errors appearing in anomalous surface trade winds, and saturate so rapidly that they primarily depend upon the seasonal cycle rather than lead time.

ENSO Previous Work

There have been some work with the same goal in the past. This paper focuses on sea surface temperature anomalies in a zonal pattern to predict ENSO. I will use time lags.

Geophysical Research Letters / Volume 50, Issue 20 / e2023GL105175

Research Letter | Open Access

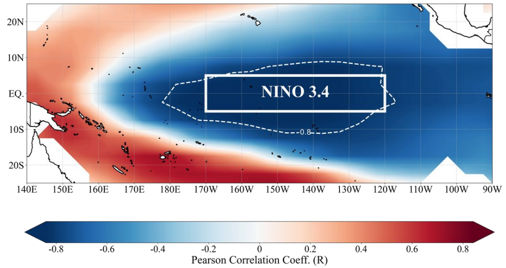
CNN-Based ENSO Forecasts With a Focus on SSTA Zonal Pattern and Physical Interpretation

Expected Outcomes

Expected outcome: accurate predictions of ONI based off sea surface temperature averages in NINO3.4 region.

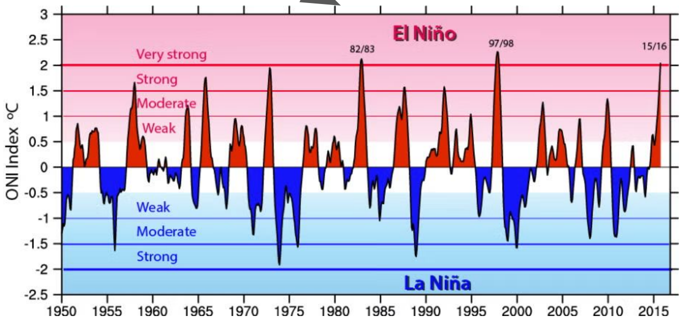
Input:

Sea surface temperature time lagged in three month intervals up to 2 years



Output:

Oceanic Nino Index (ONI)



Datasets Used

CESM

- Ensemble of 40 climate models from around the world
- Use for training and validation data (9 + 1 ensemble members)
- Variables
  - Sea Surface Temperature
  - Quarter degree grid
  - 1959 to 2015
  - NetCDF file

```
sst_test_data.SST
```

xarray.DataArray 'SST' (**ensemble_member**: 10, **time**: 673, **lat**: 22, **lon**: 46)

ERA5

- Reanalysis product (Observations + Model)
  - Proxy for observations
- Use for testing data
- Variables
  - Sea Surface Temperature
  - Eighth degree grid
  - 1959 to 2021
  - NetCDF file

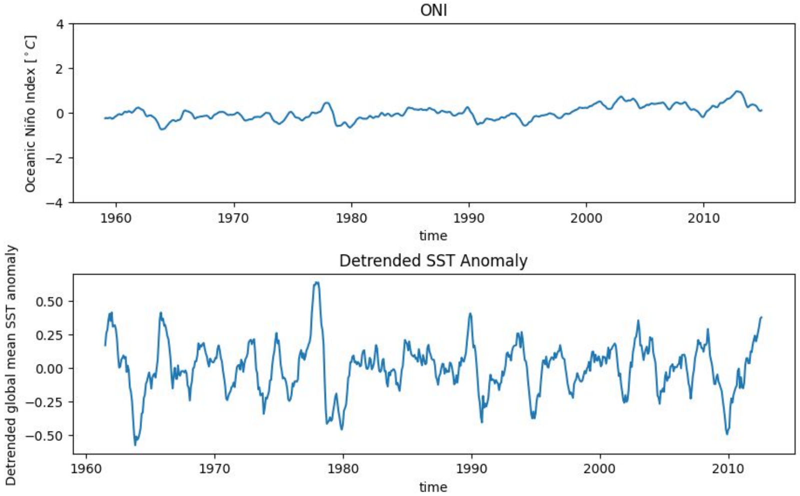
```
era
```

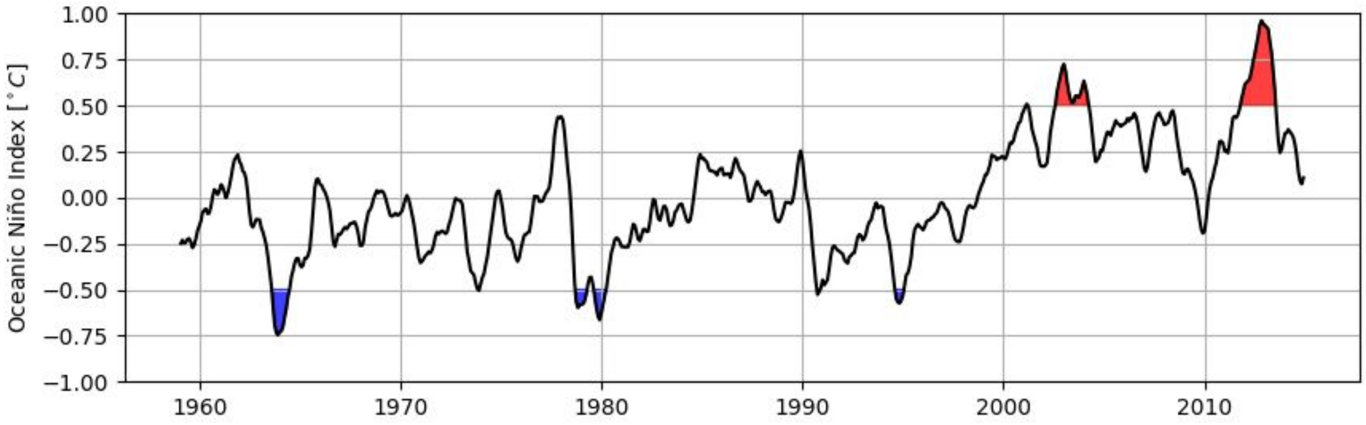
xarray.Dataset

▸ Dimensions: (**time**: 756, **lat**: 180, **lon**: 360)

Data Exploration

Example time series of one ensemble member from CESM. I have plotted ONI and detrended SST anomaly on the right and ONI with shaded regions for El Nino and La Nina on the bottom.





Data Cleansing

Because data comes from CESM and ERA5, they contain mostly clean and accurate data already. However, there are a few steps to make our computation easier:

- Convert time to a DateTime variable
- Convert from Kelvin to Celsius
- Add attributes to SST variable

```
# Extract the time variable
time_var = enso['time']

# Convert the numeric time values to datetime objects
time_units = time_var.attrs['units']
calendar = time_var.attrs.get('calendar', 'noleap')
time_dates = cftime.num2date(time_var[:], units=time_units, calendar=calendar)

# Convert cftime.DatetimeNoLeap to pandas Timestamp
def cftime_to_datetime(cftime_obj):
    return pd.Timestamp(cftime_obj.isoformat())

# Apply the conversion
time_dates_standard = [cftime_to_datetime(t) for t in time_dates]
enso['time'] = ('time', time_dates_standard)
enso = enso.sel(time=slice('1959-01-01','2021-12-01'))

enso['SST']=(enso['SST']-273.15)  # Use the formula to get from C to F.

# Add some attributes as name:value pairs.
enso['SST'].attrs['long_name']='2m temperature'
enso['SST'].attrs['units']='Fahrenheit'
enso['SST'].attrs['formula']='(K-273.15) x 1.8 + 32'
```

Data Preparation

Function to apply data cleaning to all ensemble members of CESM as well as ERA5

```
def get_climatology_data(fld, run):
    fn = f'/content/drive/MyDrive/LargeEnsemble/LargeEnsemble/ENSO_{fld}/ENSO_{fld}/ENSO_Project_{fld}_{run}p1f1.nc'
    enso = xr.open_dataset(fn, decode_times=False)

    time_var = enso['time']
    time_units = time_var.attrs['units']
    calendar = time_var.attrs.get('calendar', 'noleap')
    time_dates = cftime.num2date(time_var[:], units=time_units, calendar=calendar)

    time_dates_standard = [cftime_to_datetime(t) for t in time_dates]
    enso['time'] = ('time', time_dates_standard)

    enso = enso.sel(time=slice('1959-01-01','2021-12-01'))

    enso_po = enso.sel(lon=slice(120, 290), lat=slice(-30, 30))

    enso_clim = enso_po.groupby("time.month").mean('time')
    enso_anoms = enso_po.groupby("time.month")-enso_clim

    weights = np.cos(np.deg2rad(enso_po.lat)); weights.name = "weights"
    global_mean = enso_anoms[fld].weighted(weights).mean(dim=['lat','lon'])
    global_mean_runningMean=global_mean.rolling(time=12*5,center=True).mean()
    enso_detrended=(global_mean-global_mean_runningMean)


    return enso, enso_po, enso_clim, enso_anoms, enso_detrended
```

Calculate ONI for each of the ensemble members

Output: Oceanic Nino Index

[90] 0s

```
oni_test_data =[]
ensemble_m = ['m1','m2', 'm3', 'm4', 'm5', 'm6', 'm7', 'm8', 'm9', 'm10']

for m in ensemble_m:
    weights = np.cos(np.deg2rad(sst_test_data.sel(ensemble_member=m).lat)); weights.name = "weights" #
    NINO34 = sst_test_data.sel(ensemble_member=m).weighted(weights).mean(dim=['lat','lon'])
    ONI= NINO34.rolling(time=3,center=True).mean() # rolling mean over 3 months
    oni_test_data.append(ONI)
```

Between 1959 and 2015, there are 673 months. So we choose Nsamples = 673 and for each ensemble and each sample, we select a time lags of 3 month intervals and its corresponding ONI value for 12 months into the future.

```
Nsamples = 673

Input = np.zeros((Nsamples,len(ensemble_m), sst_test_data.SST.shape[2], sst_test_data.S
Output = np.zeros((Nsamples, len(ensemble_m), 12))
persistence = np.zeros((Nsamples,len(ensemble_m),12))

m_num=0
for m in ensemble_m:
  for i in range(Nsamples):
      # Randomly Select time index
      t0 = np.random.randint(3, sst_test_data.sel(ensemble_member=m).SST.shape[0]-13)

      # Extract past SST anomalies from different time lags (t-3 months,
      # t-6 months, etc.) as our input
      Input[i,m_num,:,:,0] = sst_test_data.sel(ensemble_member=m).SST[t0]
      Input[i,m_num,:,:,1] = sst_test_data.sel(ensemble_member=m).SST[t0-3]
      Input[i,m_num,:,:,2] = sst_test_data.sel(ensemble_member=m).SST[t0-6]
      Input[i,m_num,:,:,3] = sst_test_data.sel(ensemble_member=m).SST[t0-9]
      Input[i,m_num,:,:,4] = sst_test_data.sel(ensemble_member=m).SST[t0-12]
      Input[i,m_num,:,:,5] = sst_test_data.sel(ensemble_member=m).SST[t0-15]
      Input[i,m_num,:,:,6] = sst_test_data.sel(ensemble_member=m).SST[t0-18]
      Input[i,m_num,:,:,7] = sst_test_data.sel(ensemble_member=m).SST[t0-21]
      Input[i,m_num,:,:,8] = sst_test_data.sel(ensemble_member=m).SST[t0-24]

      # Extract future ONI values as our output
      Output[i,m_num,:] = oni_test_data[m_num].SST.isel(time=slice(t0+1,t0+13))

  m_num+=1
```

Challenges

- Using ERA5 as test data is the best way to accurate test our model on observations, however because it’s a reanalysis product, it may not be able to test the model accurately since it incorporates climate models.
- To avoid high computation costs, I use data on a quarter and eighth degree grid, but higher resolution data would yield better results
- There are many confounding variables for predicting ENSO, and here we use one single metric (time lag)

Works Cited

- https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2023GL105175
- https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/lanina/enso_evolution-status-fcsts-web.pdf
- https://www.ncei.noaa.gov/access/monitoring/enso/