



CLIENT PRESENTATION

Our Team



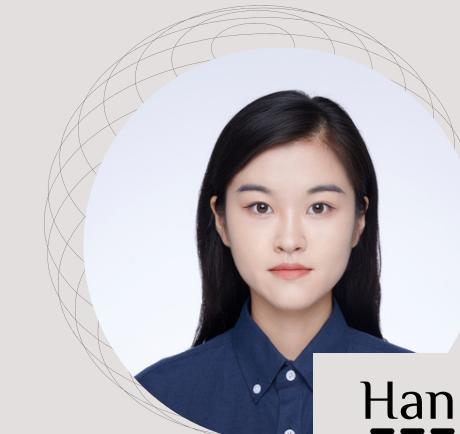
Qian Zha
Researcher



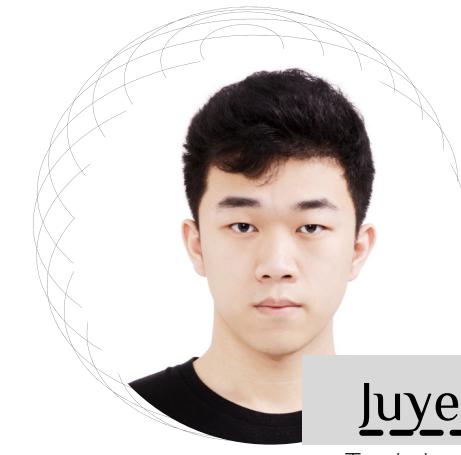
Xin Zou
Researchers, Creative
Designer



Jianglan Shi
Researcher



Han Wang
Project Manager,
Researcher



Juye Xiao
Tech Lead, Data Scientist



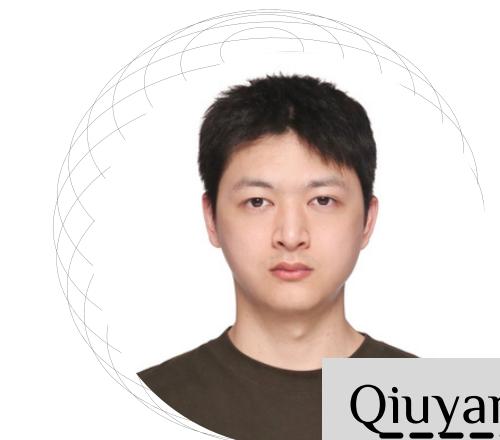
Jade Sui
Data Scientist / Engineer,
Communication Specialist



Jingkai Ma
Data Scientist



Weiqiang Guo
Data Scientist



Qiuyang Shi
Data Scientist

Agenda

Part 1. Project Backgrounds

Part 2. Exploratory Data Analysis

- Data Selection
- Heat map Visualization

Part 3. Time-series Decomposition Analysis

- Methodology Overview
- Processes and Steps

Part 4. Results and interpretation

- Results table - RMSE & R²
- Result table - Systematic difference (“delta”) identification

Part 5. Conclusions

- Conclusions and recommendations

Part 1.Project Backgrounds

- Project goal and overview
 - Data source introduction and explanation
-

Data Introduction

Source of Systematic Difference

MODIS

How was MODIS data collected ?

The MODIS instrument is operating on both the Terra and Aqua spacecraft. The **Land Surface Temperature (LST)** and Emissivity daily data are retrieved at 1km pixels by the generalized split-window algorithm and at 6km grids by the day/night algorithm.

NEX-GDDP

How was NEX data collected ?

The NEX-GDDP dataset is comprised of downscaled climate Scenarios for the globe that are derived from the General Circulation Model (GCM) runs conducted under the CMIP5 and across two of the four greenhouse gas emissions scenarios. We can retrieve the Daily **Near-Surface Air Temperature data** from NEX datasets.

SSP 126

SSP126 is based on the SSP1 assumption: "Consumption is oriented towards minimizing material resource and energy usage."



SSP 585

SSP585 is based on the SSP5(Fossil-fueled Development) assumptions : "High economic growth is combined with material intensive production and consumption patterns and a strong reliance on abundant fossil fuel resources.. "

Project Overview

Objectives:

Visually and numerically Identify the systematic difference between Land Surface Temperature (LST) from MODIS and Near Surface Temperature (TAS) from NEX models in different SSP scenarios across diverse regions across the world.

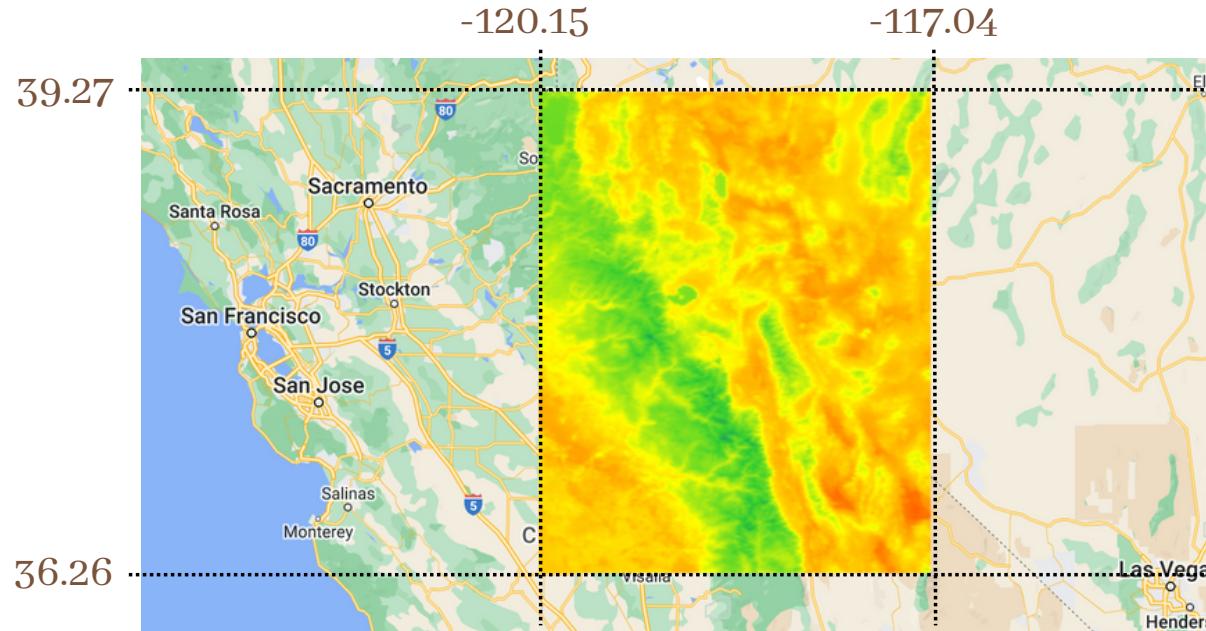
Results:

- Visualizations
 - Cross-sectional : Heat Maps at pixel level
 - Time-series: Time Series Plots for 6 years (2015 - 2020)
- Identification
 - Prediction error: RMSE; Goodness of fit: R^2
 - Difference Estimation: "Delta" at 95% Confidence Interval

Part 2.Exploratory Data Analysis(EDA)

- Data Selection - Region of Interest and Analysis Period
 - Visual Exploration - Heat Maps Visualization
-

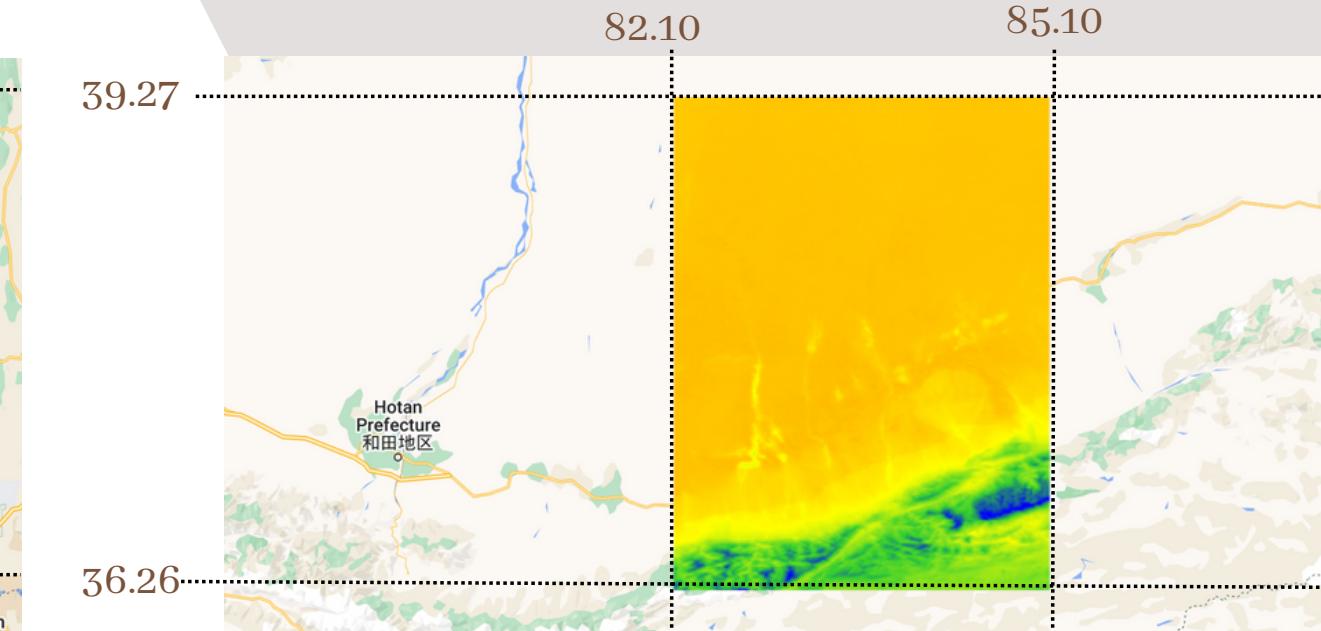
Regions



Sierra National Forest, Unite State

Longitudes and Latitudes : [-120.15,36.26,-117.04,39.27]

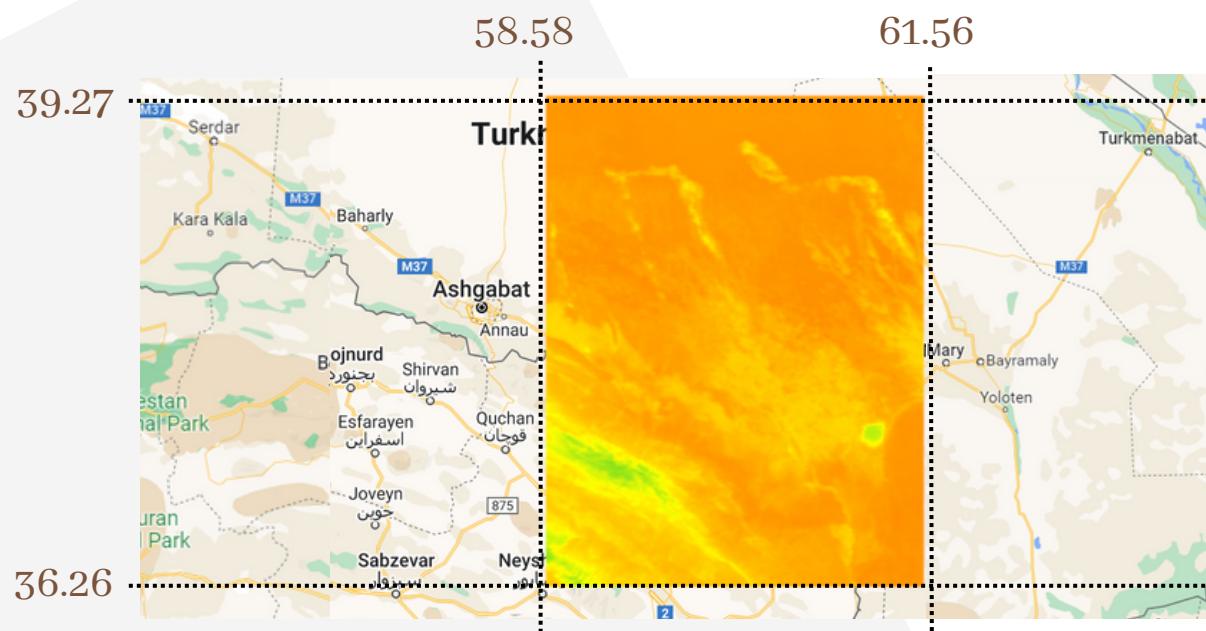
Area size: 300km * 300km



Xin Jiang, China

Longitudes and Latitudes : [82.10,36.26,85.10,39.27]

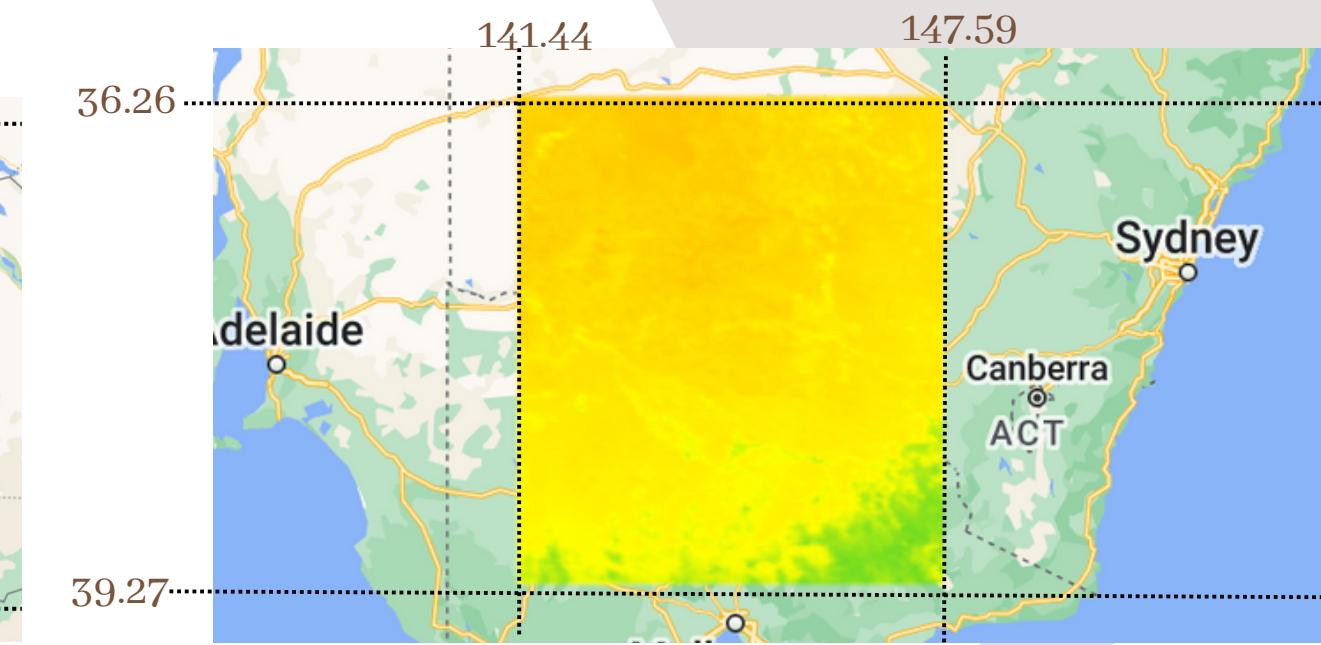
Area size: 300km * 300km



Mary, Turkmenistan

Longitudes and Latitudes : [61.56,36.26,58.58,39.27]

Area size: 300km * 300km



New south wales, Australia

Longitudes and Latitudes : [141.44,-37.66, 147.59,-31.66]

Area size: 600km * 600km

Region Selection Rationale

Project Goal:

- Identifying the variations in systematic differences between MODIS and NEX models **across four different regions.**

Motivation:

- Same latitude areas may have similar climate which minimize the influence of projection angle to our test
- Having the Southern Hemisphere region as an reflection group to increase the universality of our result

Approaches:

- Obtaining coordinates from google map, and library-related package(RClimChange) in Python or R
- Input desired coordinates (roi) in MODIS NEX
- Input the model name, years, scenario, variable

Example:

```
gcm_download_data(location=getwd(),
                   scenario='ssp126',
                   variable='tas',
                   years=2020,
                   roi=c(82.104994, 85.105204, 36.25770090268727, 39.27066973350914),
                   method='curl',
                   model=c('CNRM-CM6-1'))
```

Date Range Selection

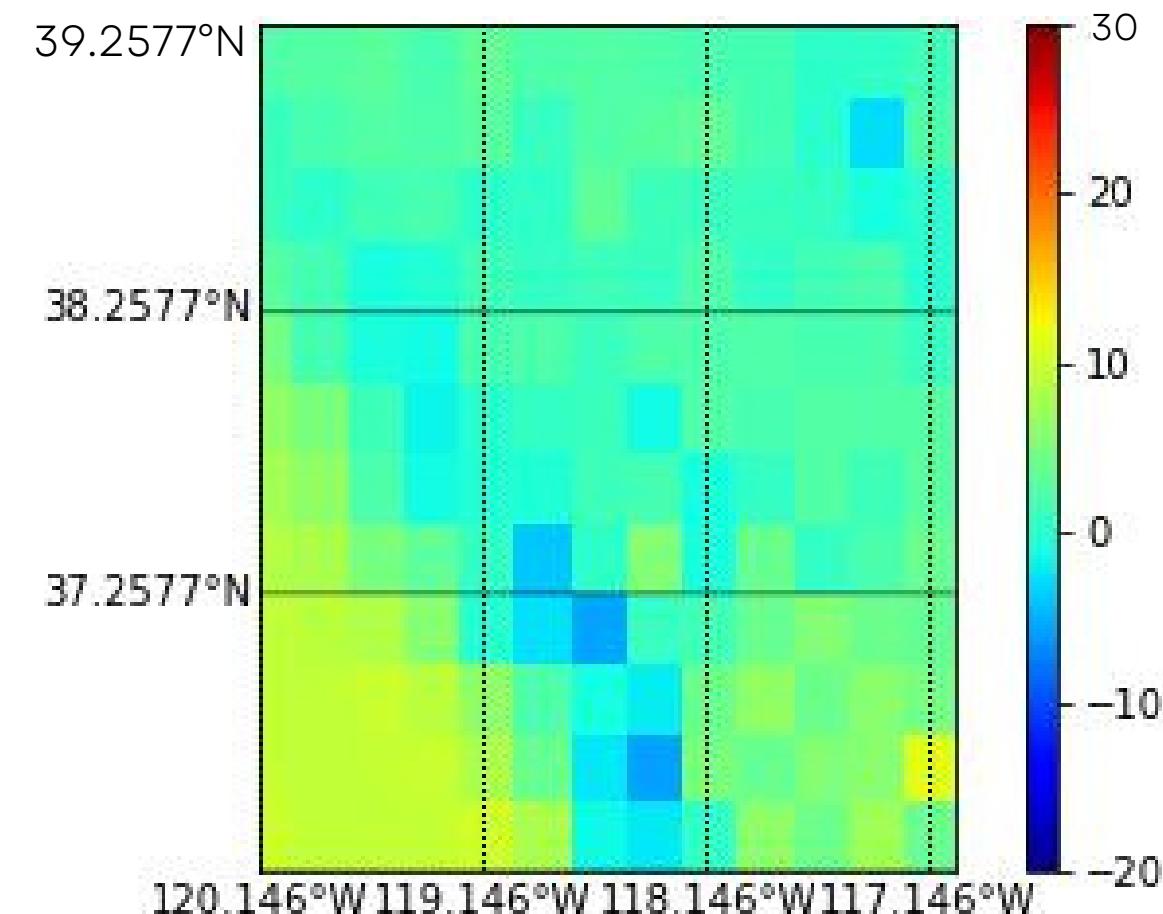
The date range selected for analysis is from 2015 - 2020.

The data for GCM NEX Models is available from 1850 but this project aims to compare across scenarios (SSP 126 vs SSP 585) so only the date range that are applicable are selected.

Model Configuration	Period/Length	Experiment(s)
CMCC-CM2-SR5 <i>Resolution : 1°x1°</i>	500 yrs	piControl (pre-industrial)
	1850-2014	historical
	2015-2100	ssp126, ssp245, ssp370, ssp585 (scenarios)
	165 yrs	abrupt-4xCO2, 1pctCO2
	310 yrs	omip1, omip2
CMCC-ESM2 <i>Resolution : 1°x1°</i>	500 yrs	piControl (pre-industrial)
	1850-2014	historical
	2015-2100	ssp126, ssp245, ssp370, ssp585, ssp534-over (scenarios)
	165 yrs	abrupt-4xCO2, 1pctCO2, 1pctCO2-bgc, 1pctCO2-rad
	310 yrs	omip1
	450 yrs	LUMIP & LS3MIP cases

Visual Explorations: Heat Maps

CA_ssp126_NorESM2_MM Near-Surface Air Temperature for Jan 2015



Y axis : Longitudes of Interested region

X axis : Latitudes of Interested region

Legend: Near Surface Temperature Scale (red color: high temperature)

Methodology Overview

The goal of heat map is to breakdown an entire region ($300\text{km} * 300\text{km}$) into small pixels ($25\text{km} * 25\text{km}$) and visualize the difference between MODIS and NEX at pixel level.

Procedures

The procedures of creating heat maps for both data sources are similar. However, there is one important extra step for MODIS data: resampling image from 1km to 25km

Step 1. Flatten 3-Dimensional monthly data (time, lon, lat) to 2-Dimensional data (lon, lat)

```
> Dimensions: (lon: 13, lat: 12, time: 2190)  
  
▼ Coordinates:  
  lon (lon) float64 -120.1 -119.9 ... -117.4 -117.1  
  lat (lat) float64 39.12 38.88 38.62 ... 36.62 36.38  
  time (time) object 2015-01-01 12:00:00 ... 2020-12...  
  
▼ Data variables:  
  tas (time, lat, lon) float32 dask.array<chunksize=(365, 12, ...)  
  
► Attributes: (0)
```

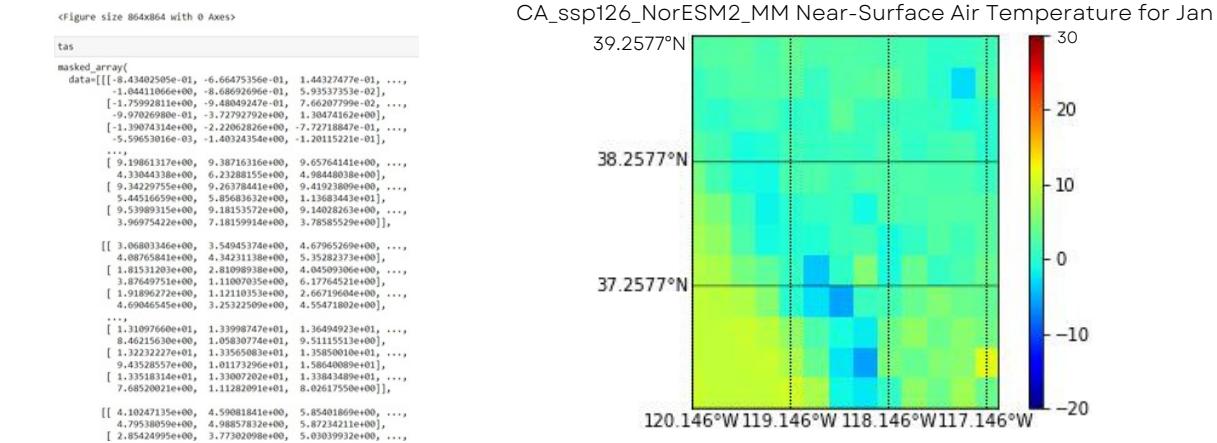
```
lon, lat=np.meshgrid(lons, lats)  
x, y=mp(lon, lat)  
plt.figure(figsize=(20, 20))
```

Step 2 . Scale temperature to degree Celsius

Step 3 . Load base-map for selected region (Example code shows Sierra National Forest, CA coordinates)

```
mp = Basemap(projection='merc',  
             llcrnrlon=-120.1456853999999,  
             llcrnrlat=36.25770090268727,  
             urcrnrlon=-117.03655454062499,  
             urcrnrlat=39.27066973350914,  
             resolution = 'i')
```

Step 4 . Map monthly temperature for each pixel onto the base map by looping through the dataset from January to December



Procedures - MODIS Heat Maps

Why resampling techniques is used ?

MODIS data are retrieved at 1km pixels while NEX data are retrieved from 25km pixels, in order to compare apple to apple. We have to **regrid MODIS data to 25km pixels** as well

What resampling method is used ?

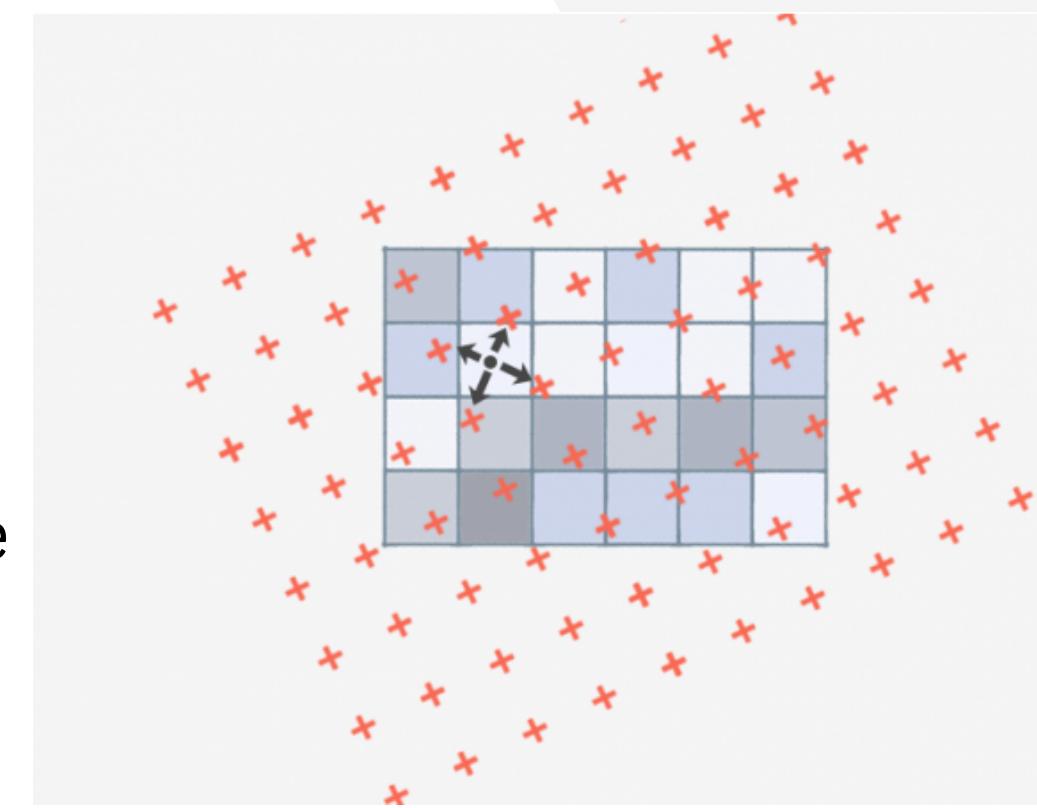
We used **bilinear interpolation** from nctoolkit package to resample the MODIS data.

Why do we use such resampling method ?

Bilinear interpolation uses **4 nearest neighbors** to generate an output surface.

Bilinear interpolation assumes the input is continuous.

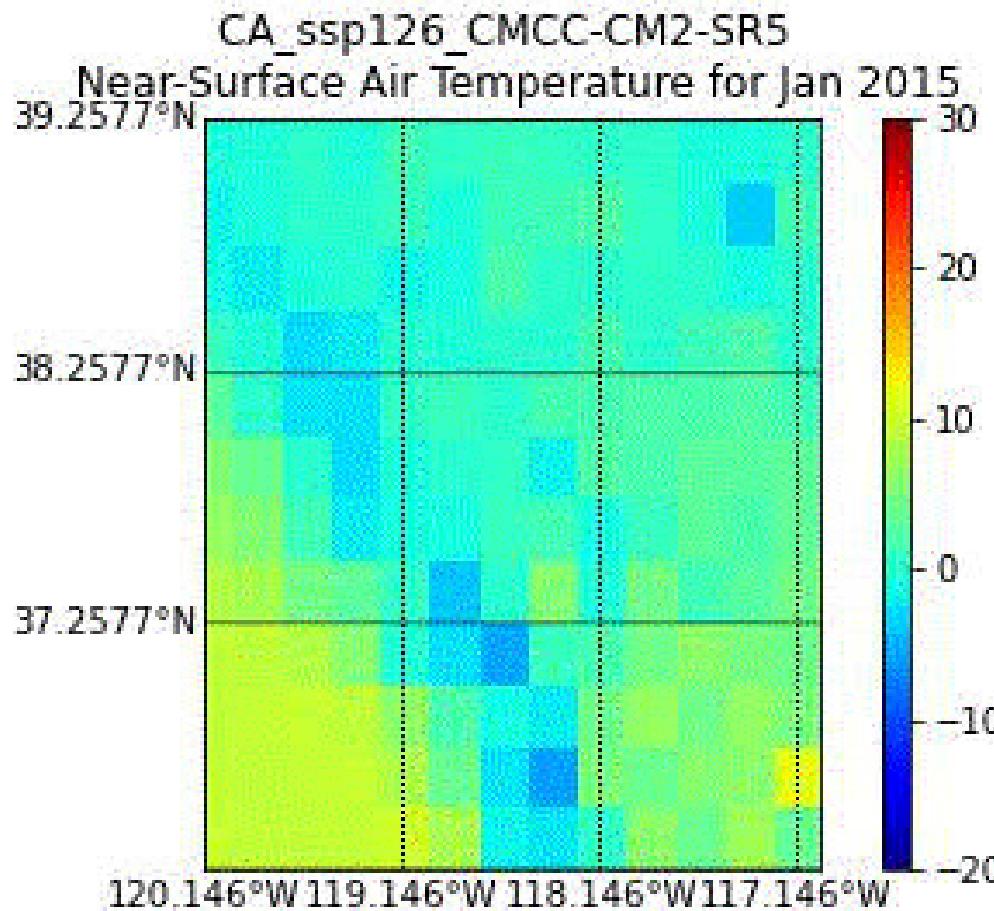
This resampling method uses a distance average to estimate with **closer cells being given higher weights**.



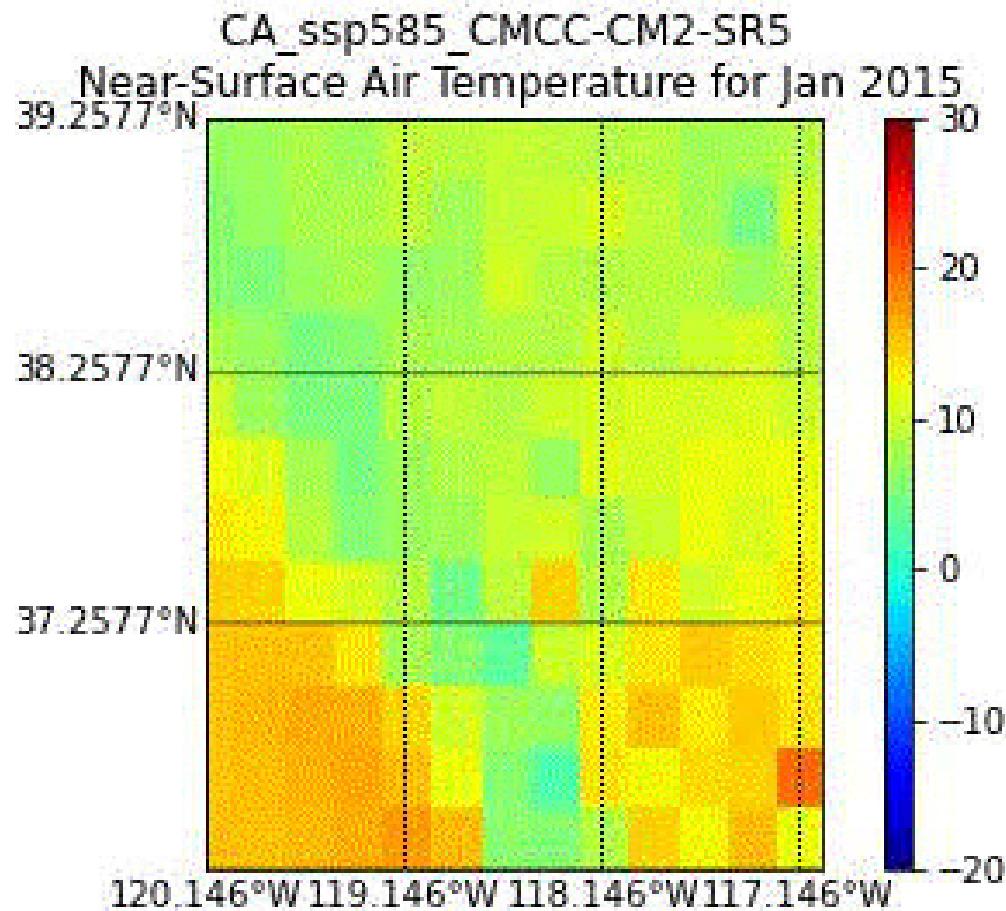
Visual Explorations: Heat Maps

NEX: CMCC-CM2-SR5

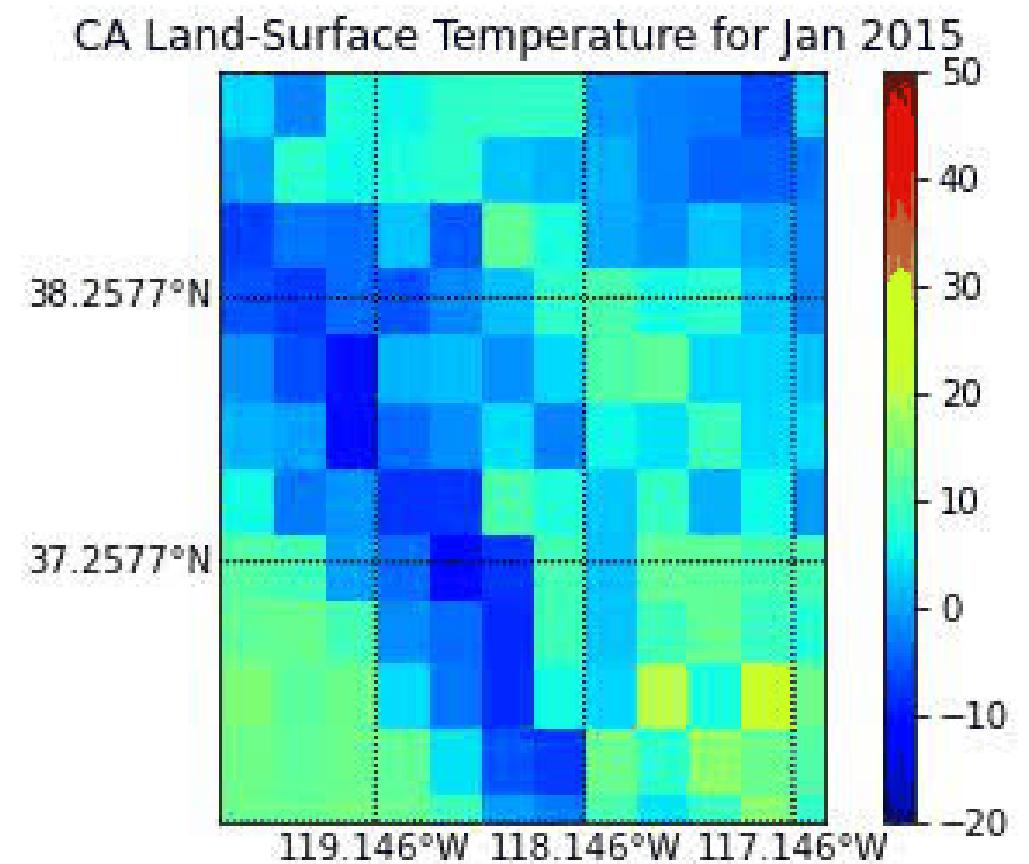
SSP 126



SSP 585



MODIS



Monthly average near surface temperature in 2015 at Sierra National Forest

Monthly average land surface temperature in 2015 at Sierra National Forest

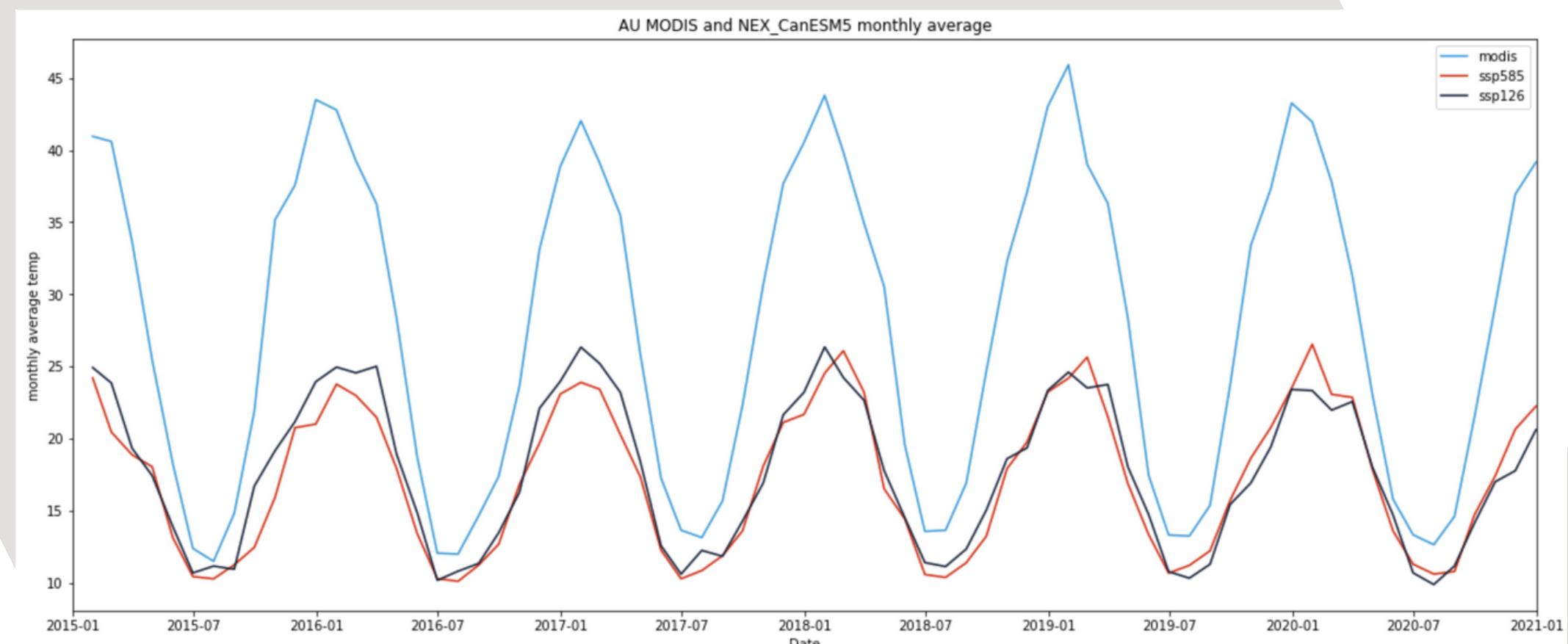
Part 3. Time-series analysis

- Methodology Overview
 - Time-series
 - Monthly average value from 2015 to 2020
 - Time-series decomposition
 - Visualize systematic difference in trends
 - Comparison between MODIS and selected NEX model
-

Time-series analysis methodology overview

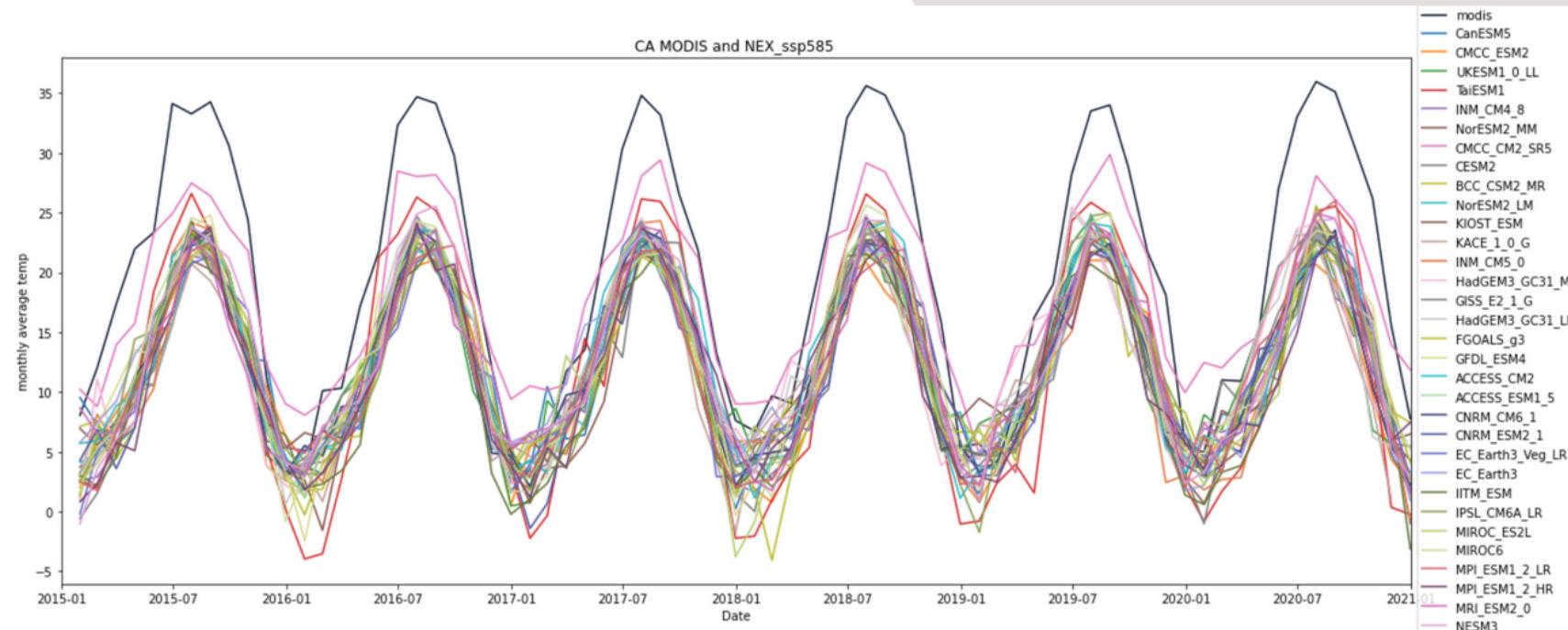
Why did we choose time series plots ?

- Compare the monthly mean temperature value between MODIS and 32 NEX models in 2 scenarios from 2015 to 2020
- Allow us to compare the difference between SSP126 and SSP585 performance against MODIS at each region
- Straightforward visualization and easy to Interpret



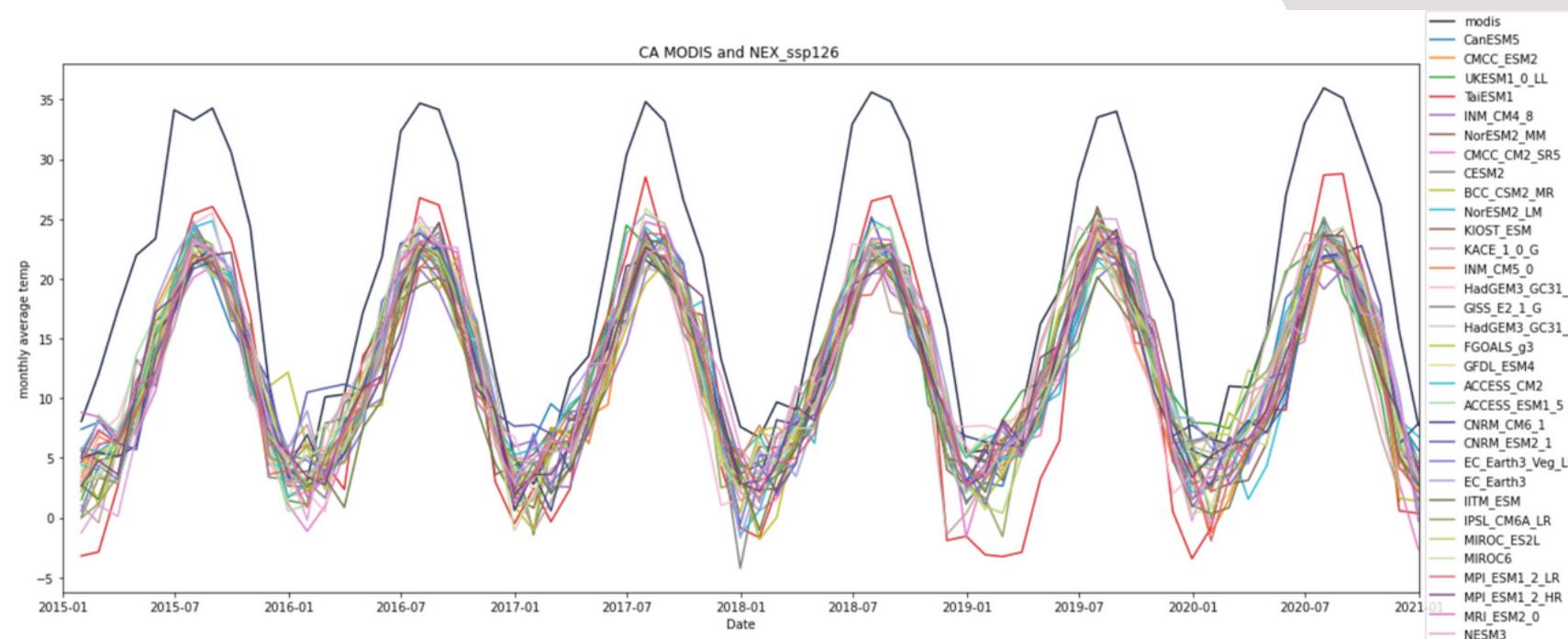
Time-series Plot

Sierra Natural Forest, CA



MODIS vs. 32 NEX models

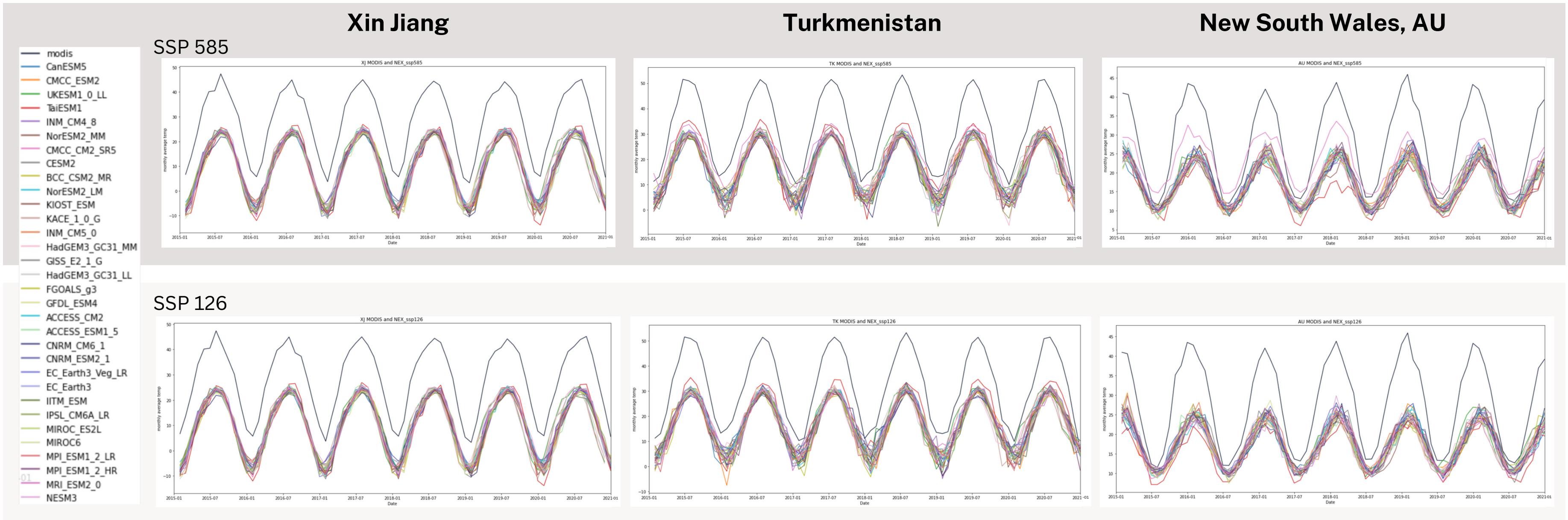
The systematic differences between near surface temperature (NEX) and land surface temperature (MODIS)



The difference between SSP 126 and SSP 585 for same NEX model is minimal compared with their gap against MODIS

Consistency in other regions

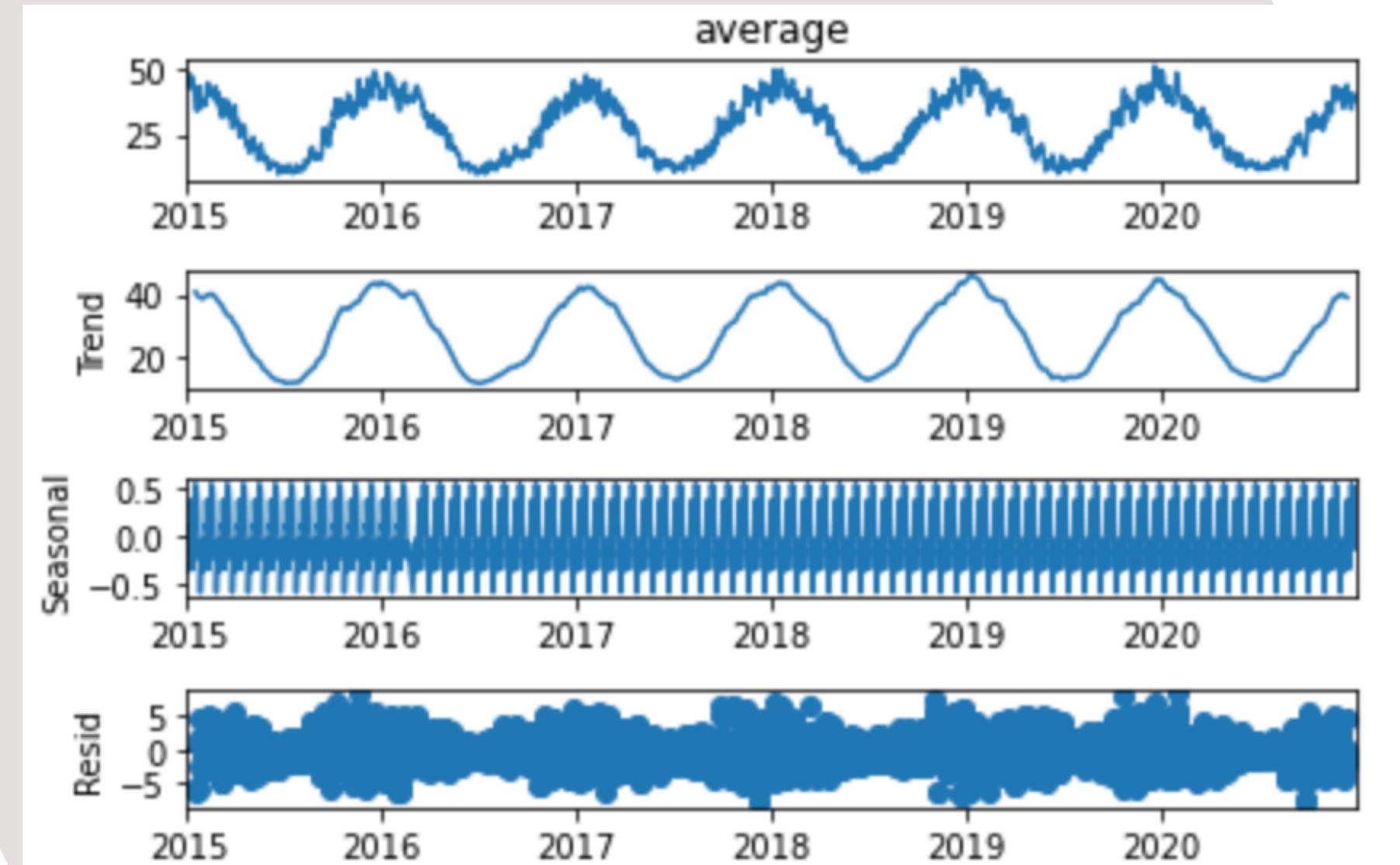
Applying the same time-series analysis across the other regions, found consistent results.



MODIS vs. 32 NEX models

Time-series decomposition methodology overview

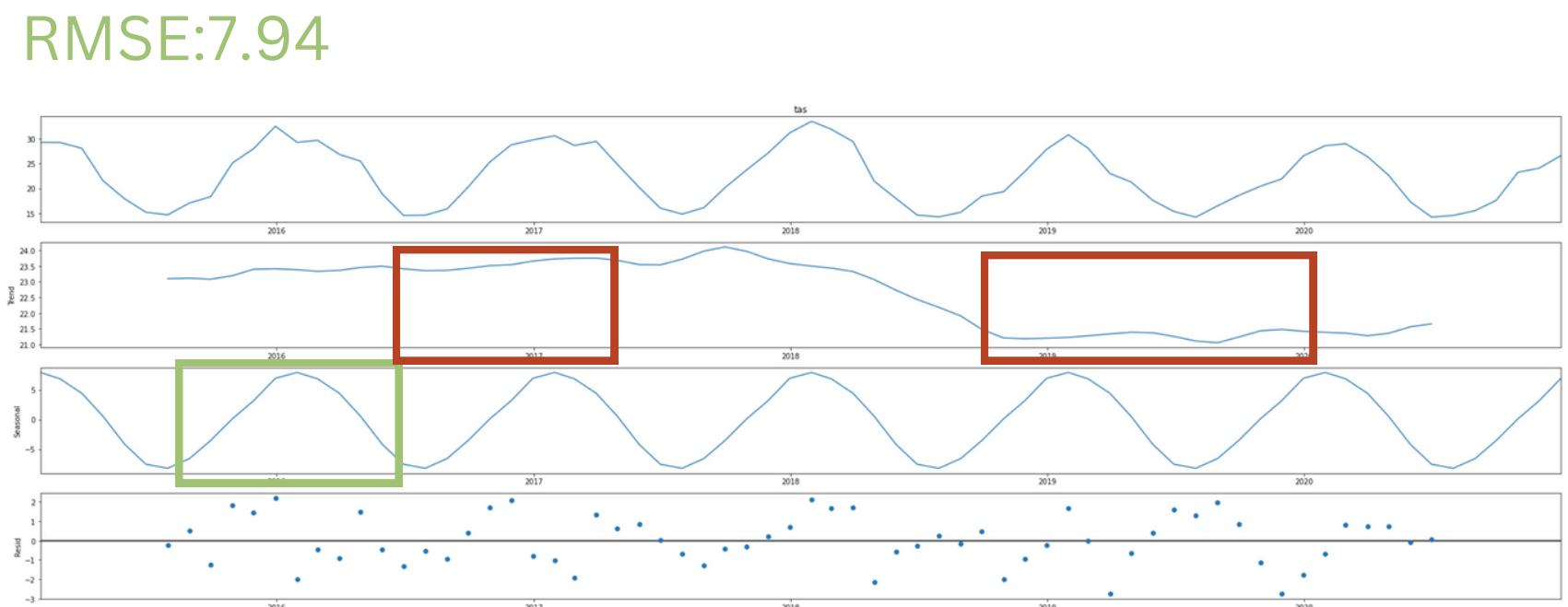
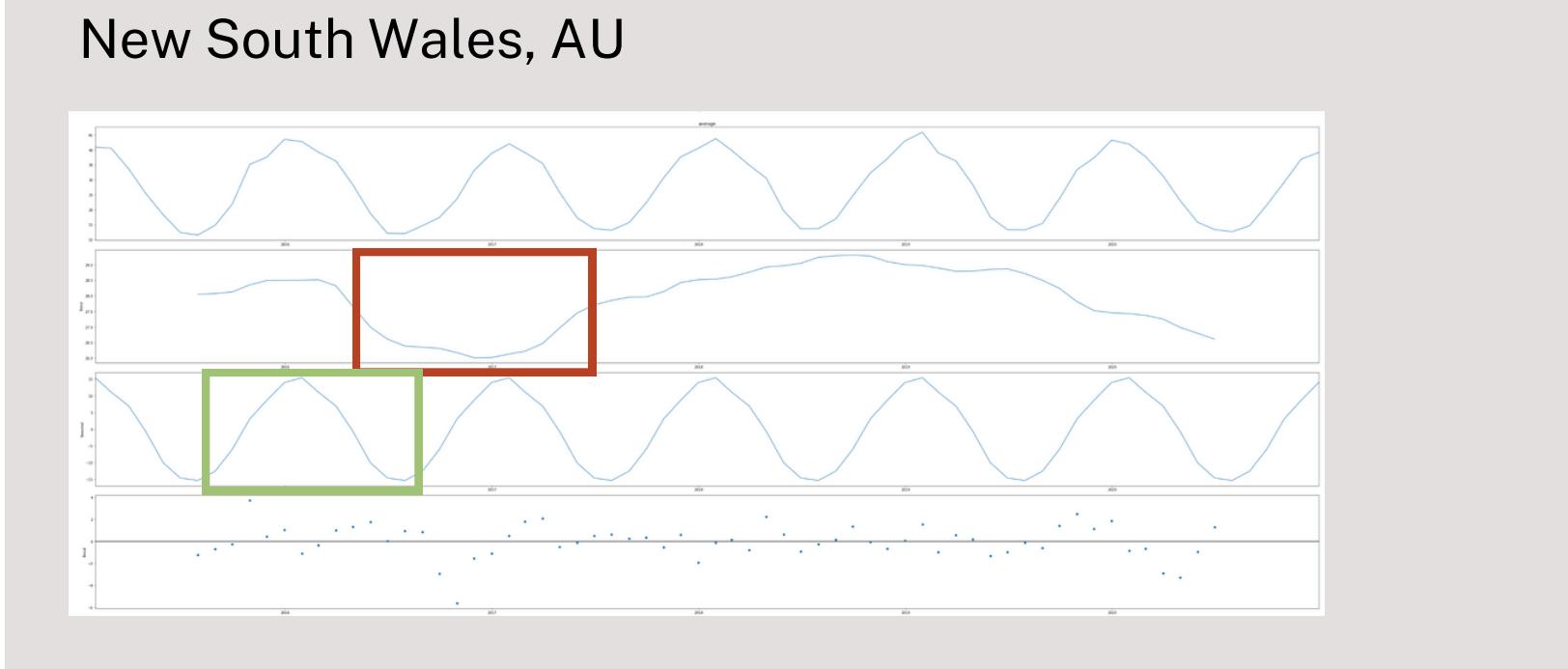
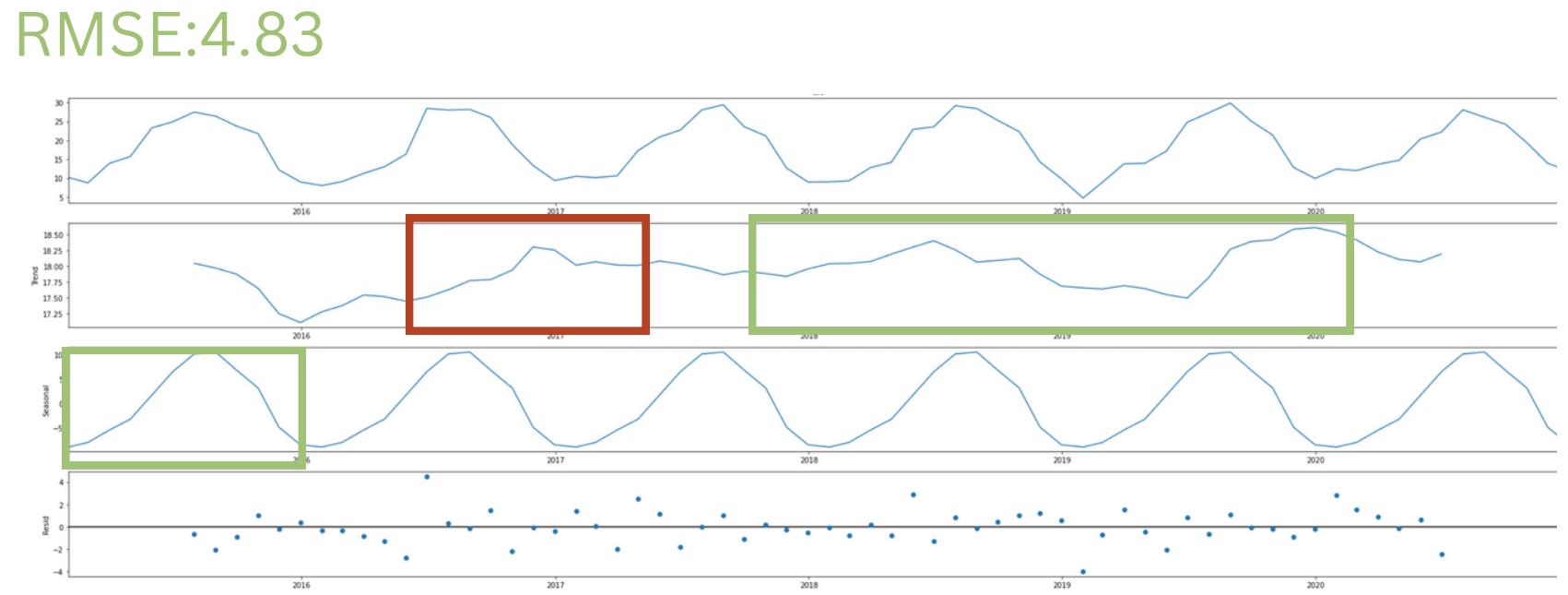
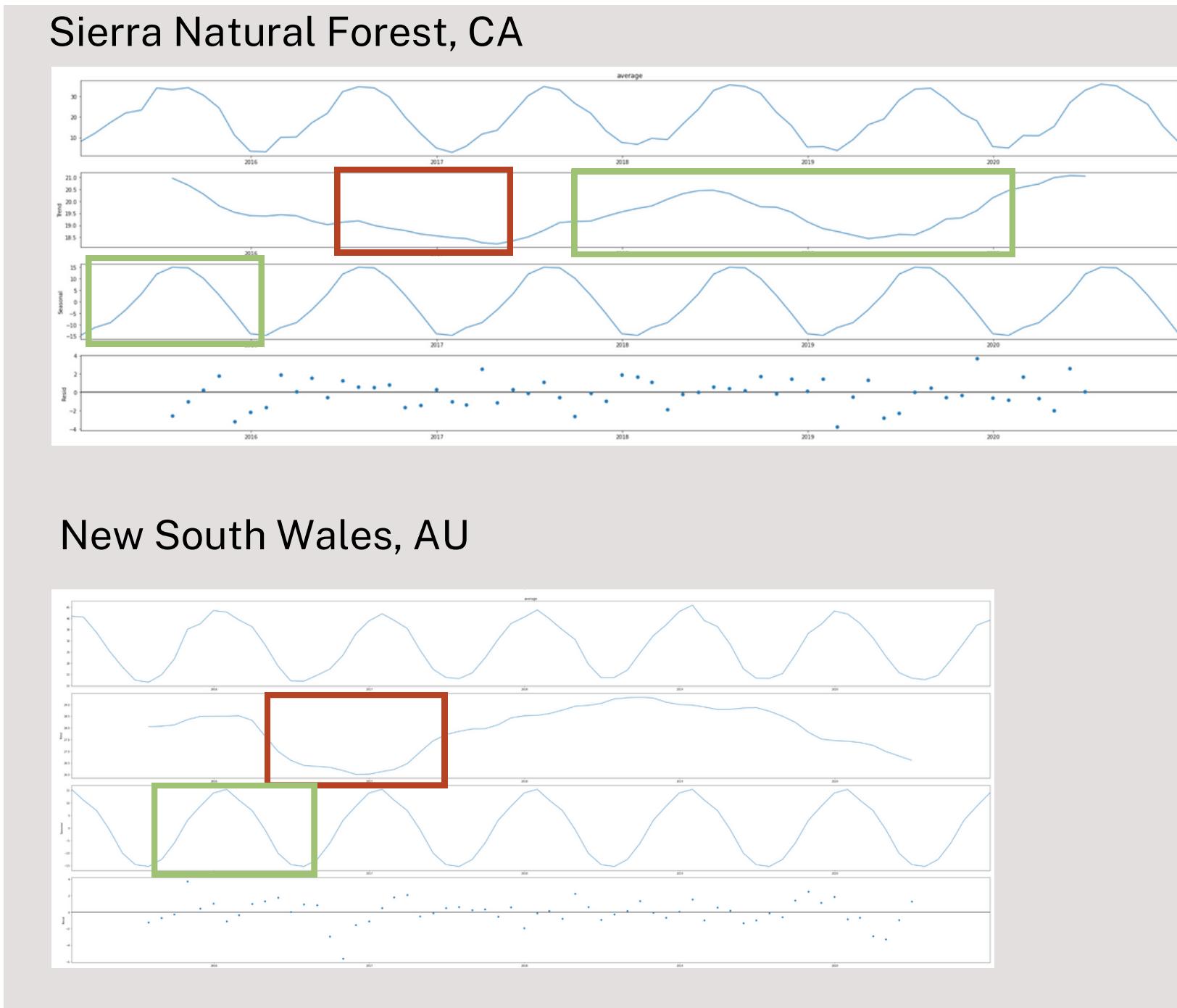
- Decompose the general pattern into trend, seasonality, residues for deep-dive analysis
- Allow us to make side by side comparison to see the performance of each model



MODIS daily land surface temperature New South Wales, AU

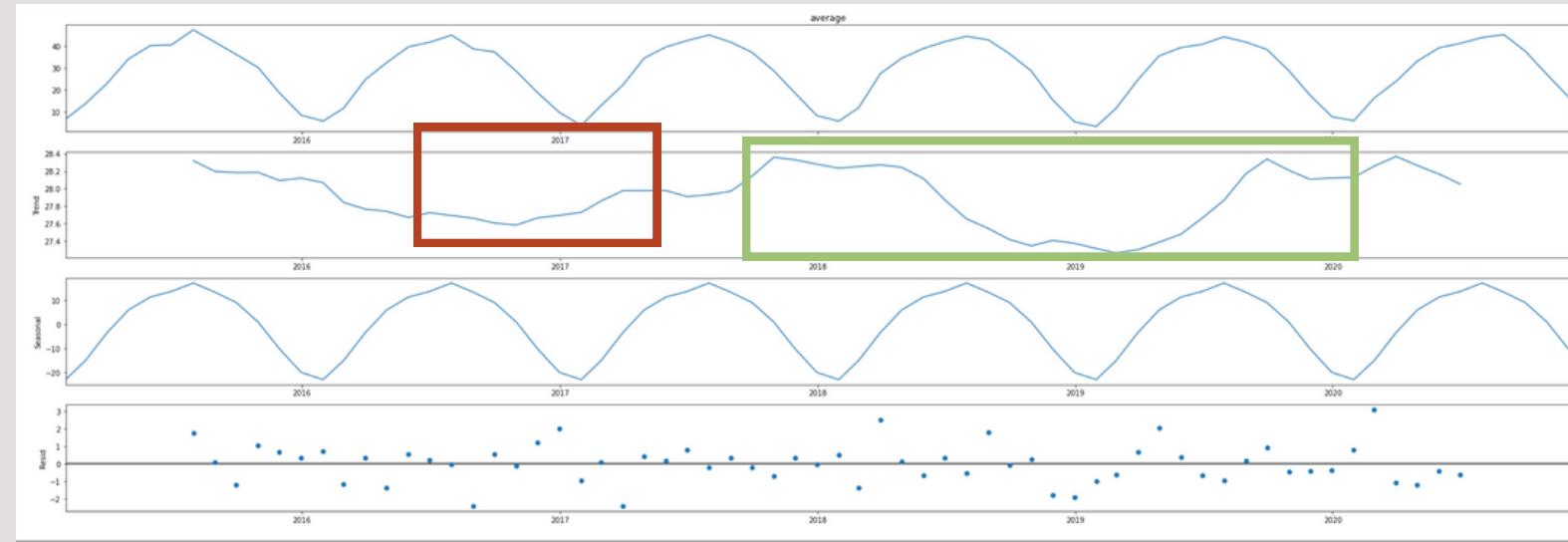
MODIS vs. CMCC-CM2-SR5

Select the NEX model CMCC-CM2-SR5 under SSP 585 scenario (lowest RMSE NEX model) and compare with MODIS side by side across all the regions. We have Identified that major difference between MODIS and CMCC-CM2-SR5 occur in the trend component.

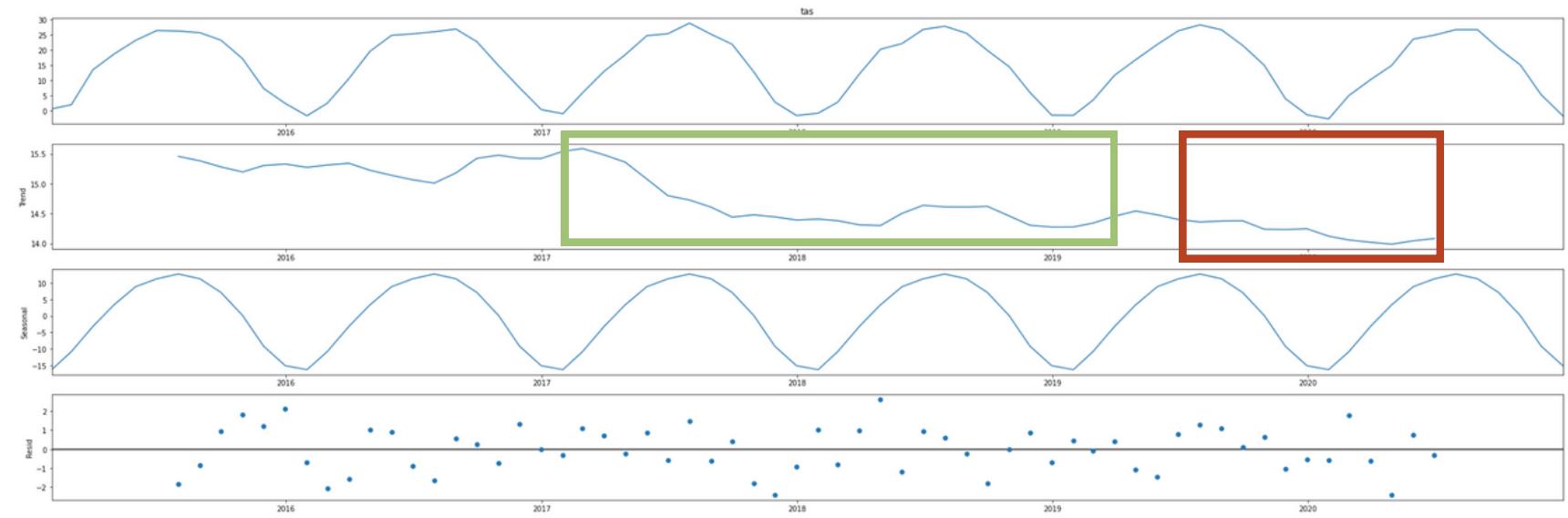


MODIS vs. CMCC-CM2-SR5

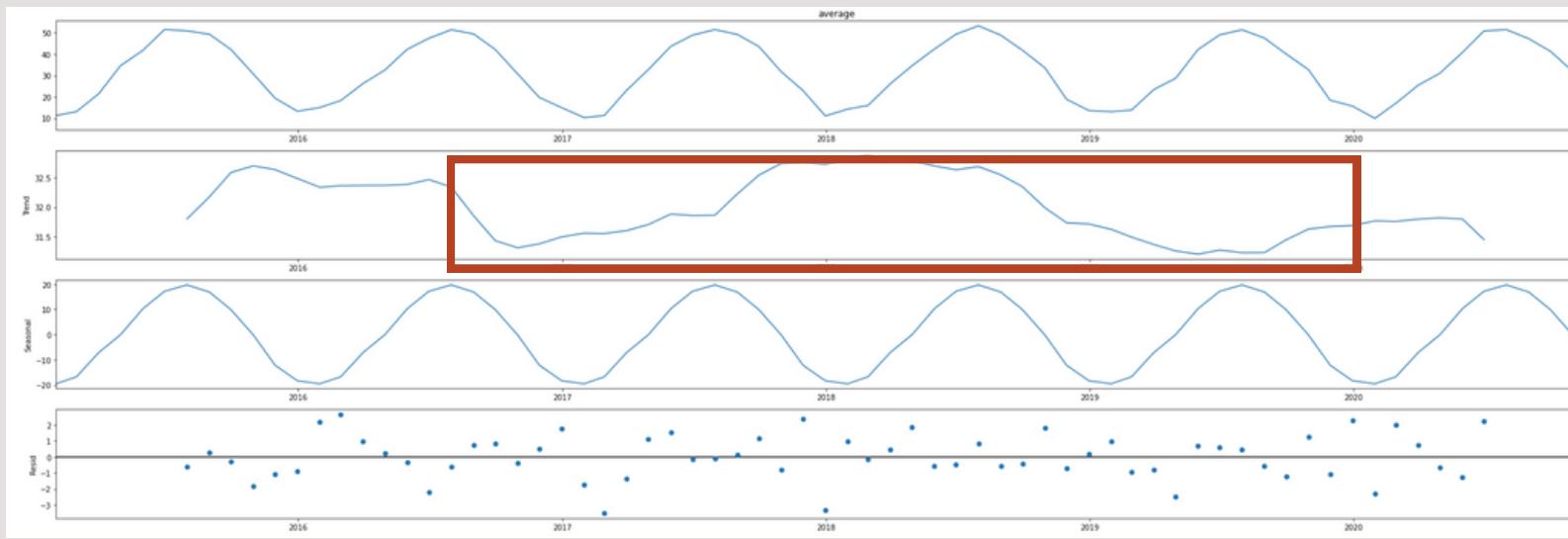
Xin Jiang



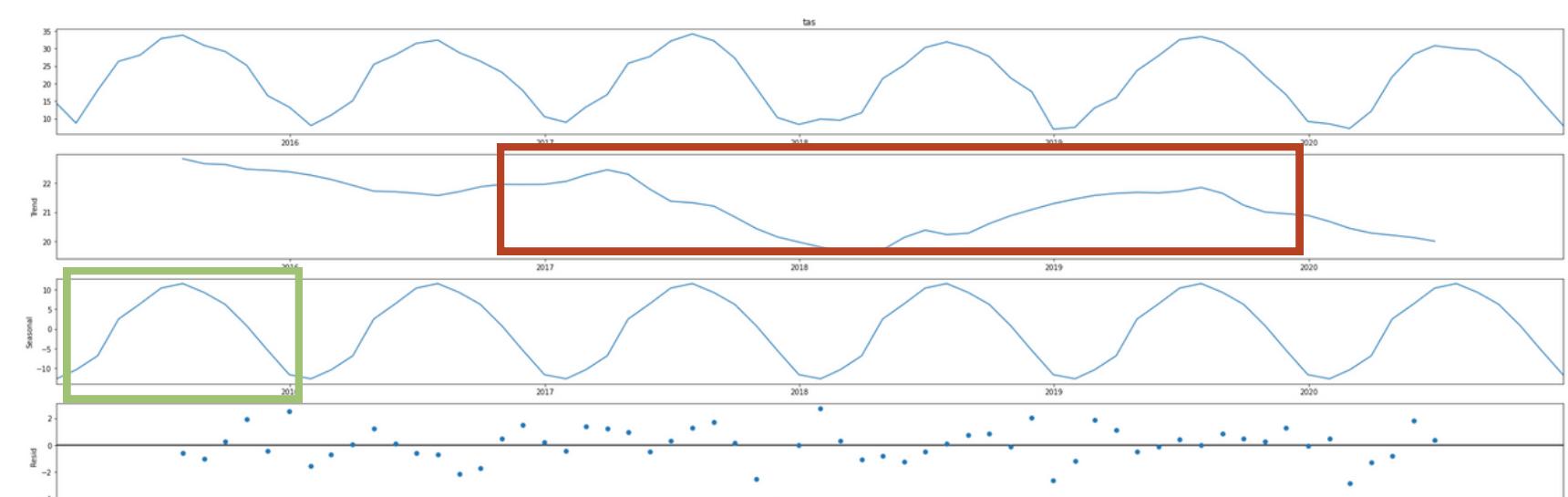
RMSE:13.73



Mary, Turkmanistan



RMSE:15.84



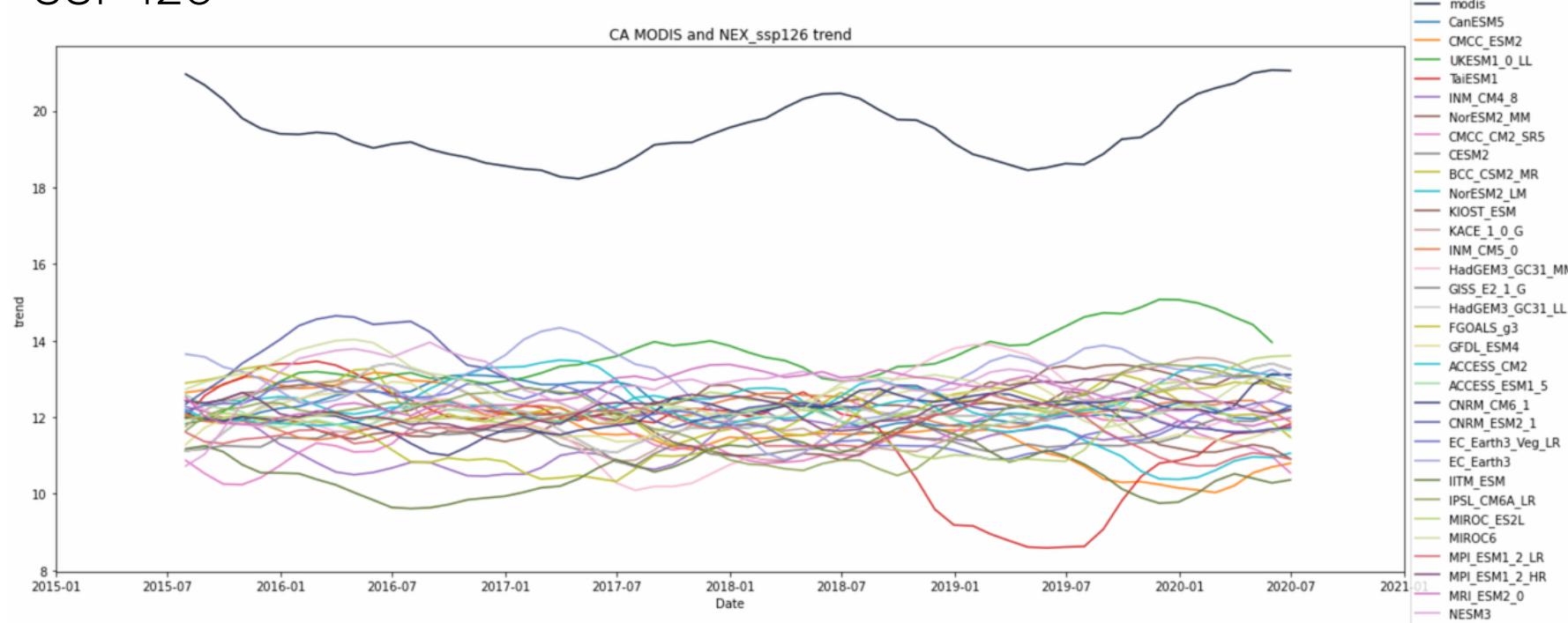
MODIS

CMCC-CM2-SR5: SSP 585

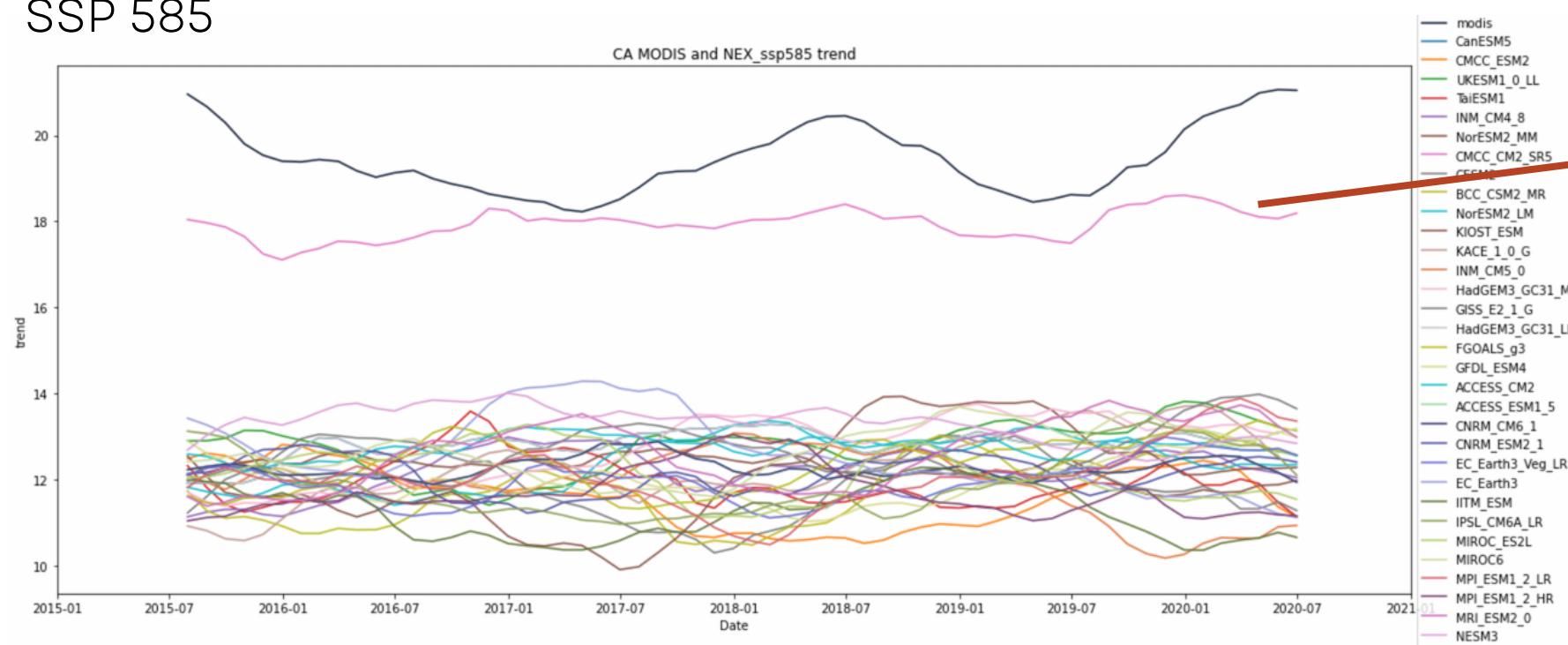
Trend components

Sierra National Forest, CA

SSP 126



SSP 585



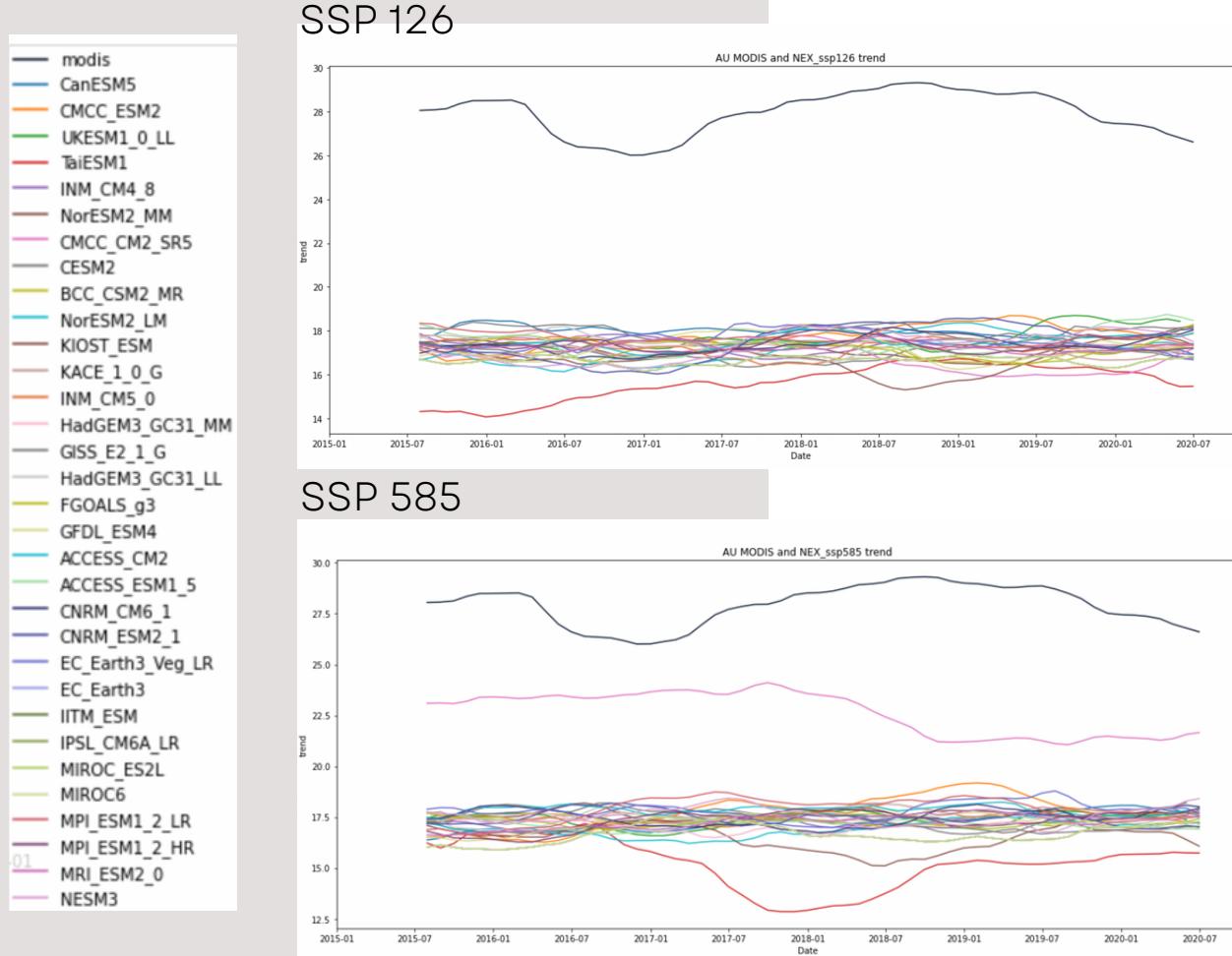
The trend components of NEX models are mostly stacks at the within an upper and lower bound. The majority of which are clustered together with only a few outliers such as CMCC-CM2-SR5.

CMCC-CM2-SR5

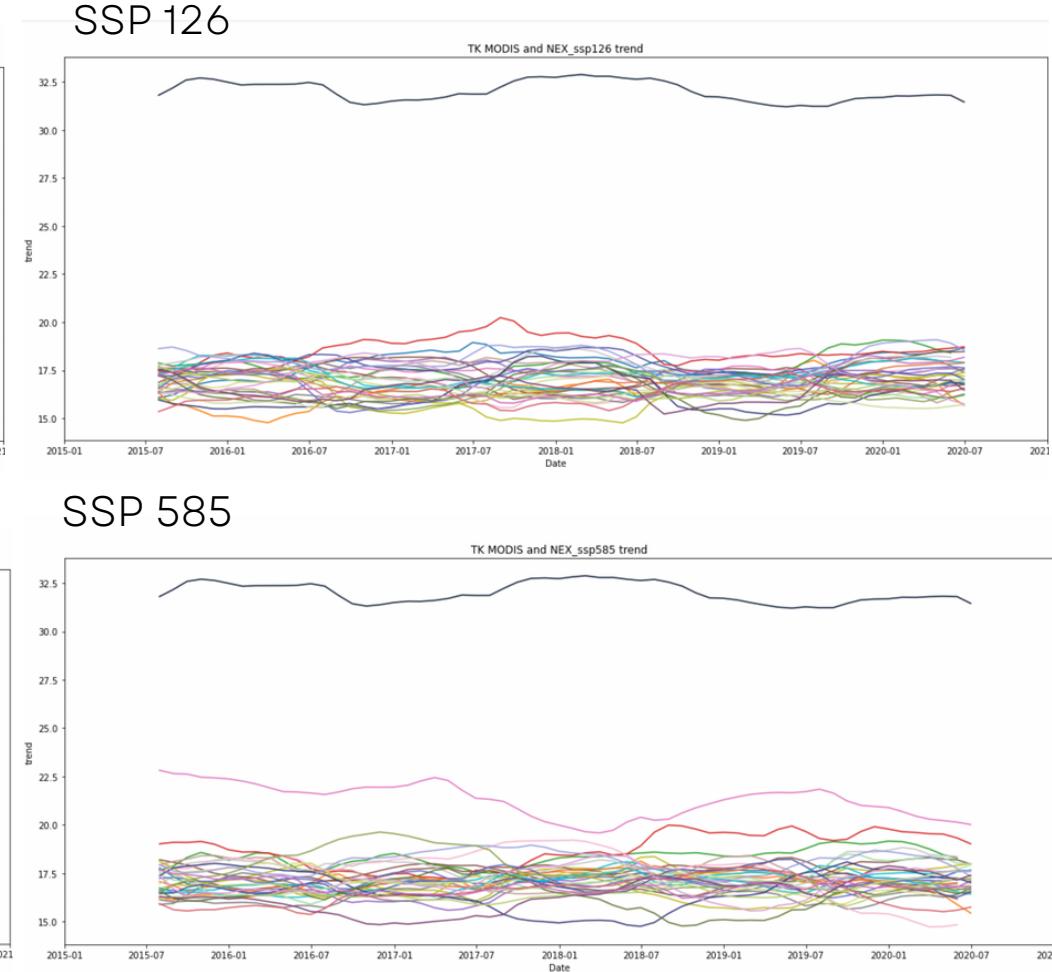
Trend components

Although the boundary of two SSP scenarios do not differ significantly, the SSP 126 tend to stack "tighter" than SSP 585, and we found it consistent across other regions

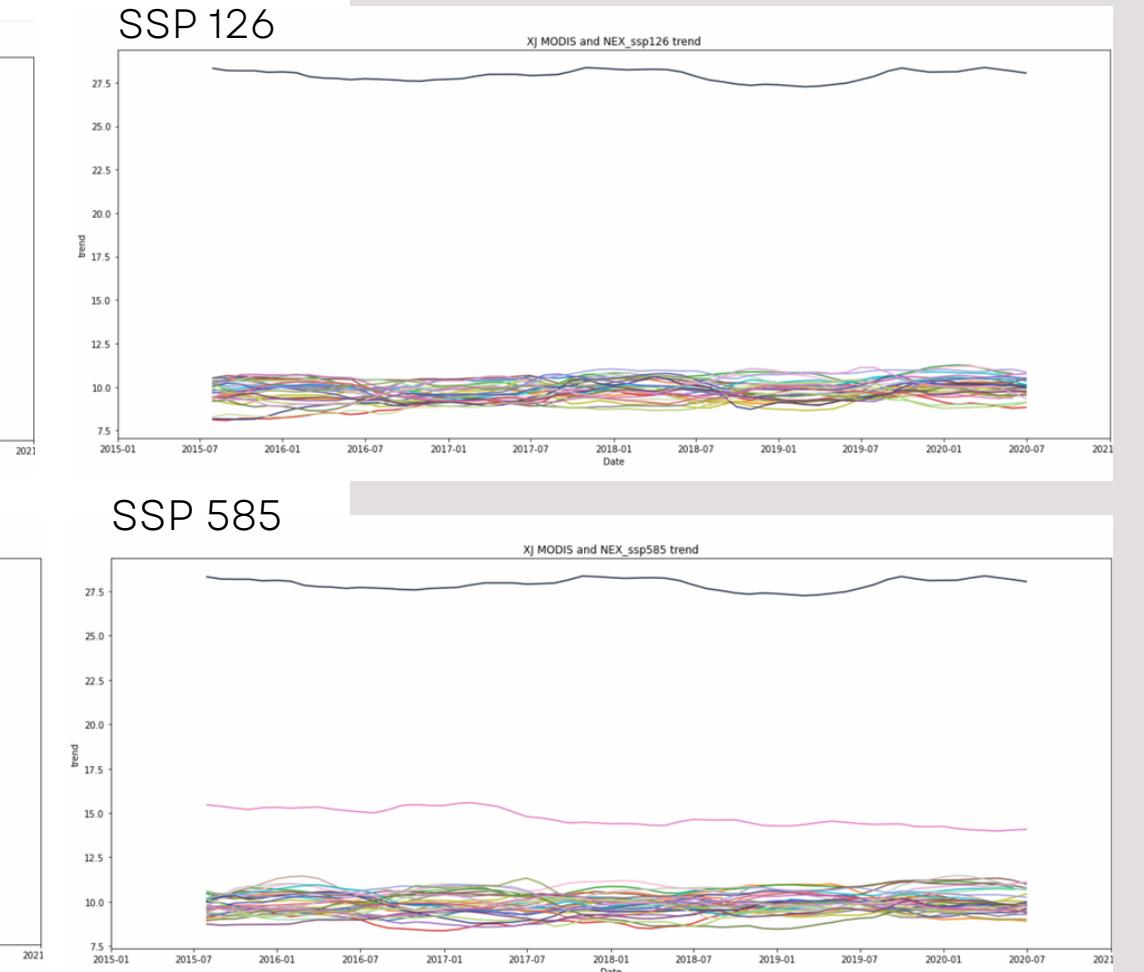
New South Wales, AU



Mary, Turkmanistan



Xin Jiang, China



SSP 126 is based on the SSP1 assumption: "Consumption is oriented towards minimizing material resource and energy usage."

SSP 585 is based on the SSP5 assumptions : "Fossil-fueled Development. The social and economic development, however, is based on an intensified exploitation of fossil fuel resources with a high percentage of coal and an energy-intensive lifestyle worldwide."

Part 4. Results and Interpretation

Project goal: identifying the variations in systematic differences between MODIS and NEX models across four different regions.

- Methodology Overview
 - Procedures
 - Results - Metrics for difference identification: RMSE, R^2 and systematic difference ("delta") estimation
-

Methodology Overview

RMSE

The root mean square error (RMSE) allows us to measure how far predicted monthly near surface temperature (NEX) are from observed monthly land surface temperature (MODIS). The RMSE result is positively correlated with the prediction error.

R²

The R² shows the goodness of fit of each NEX model in percentage, there is a positive relationship between R² and the proportion of variance in the MODIS that can be explained by the NEX model.

"Delta"

Estimating the systematic difference between land surface temperature and near surface temperature at 95% confidence level.

RMSE Calculation

Step1. Return the RMSE with monthly value

```
#calculate RMSE
TK_ssp585_NESM3_monthly_rmse = sqrt(mean_squared_error(TK_Monthly_data_month, TK_ssp585_NESM3_monthly_avg_temp))
TK_ssp585_NESM3_monthly_rmse
```

↳ 15.748301635369108

R[^]2 Calculation

Step 1. Perform linear regression using NEX model as predictor variable and MODIS as outcome variable

```
TK_ssp585_NESM3_model = LinearRegression()
x = TK_ssp585_NESM3_monthly_avg_temp.values.reshape(-1, 1)
y = TK_Monthly_data_month.values.reshape(-1, 1)
```

Step 2. Return the R[^]2

```
TK_ssp585_NESM3_model.fit(x, y)
TK_ssp585_NESM3_r_squared =TK_ssp585_NESM3_model.score(x, y)
TK_ssp585_NESM3_r_squared
```

↳ 0.9183710762349819

"Delta" Estimation Steps

Step1. Aggregate the monthly means by separating the annual data into 4 seasons (every 3 months) over the 6-years period.

```
[ ] month = TK_ssp585_NESM3_monthly_avg_temp.index.month
TK_ssp585_NESM3_winter = pd.concat([TK_ssp585_NESM3_monthly_avg_temp[month == 1], TK_ssp585_NESM3_monthly_avg_temp[month == 2],TK_ssp585_NESM3_monthly_avg_temp[month == 12]])
TK_ssp585_NESM3_spring = pd.concat([TK_ssp585_NESM3_monthly_avg_temp[month == 3], TK_ssp585_NESM3_monthly_avg_temp[month == 4],TK_ssp585_NESM3_monthly_avg_temp[month == 5]])
TK_ssp585_NESM3_summer = pd.concat([TK_ssp585_NESM3_monthly_avg_temp[month == 6], TK_ssp585_NESM3_monthly_avg_temp[month == 7],TK_ssp585_NESM3_monthly_avg_temp[month == 8]])
TK_ssp585_NESM3_fall = pd.concat([TK_ssp585_NESM3_monthly_avg_temp[month == 9], TK_ssp585_NESM3_monthly_avg_temp[month == 10],TK_ssp585_NESM3_monthly_avg_temp[month == 11]])
TK_ssp585_NESM3_winter.sort_index()
TK_ssp585_NESM3_spring.sort_index()
TK_ssp585_NESM3_summer.sort_index()
TK_ssp585_NESM3_fall.sort_index()

time
2015-09-30    24.539616
2015-10-31    21.349688
2015-11-30     9.068030
2016-09-30    24.263397
2016-10-31    19.785851
2016-11-30    12.415395
2017-09-30    24.264217
2017-10-31    17.208358
2017-11-30    12.118357
2018-09-30    24.771110
2018-10-31    16.649083
2018-11-30    11.280638
2019-09-30    27.423459
2019-10-31    19.837588
2019-11-30    10.343564
2020-09-30    24.374521
2020-10-31    21.212649
2020-11-30     8.999966
Name: tsc, dtype: float64
```

"Delta" Estimation Steps

Step1. Perform linear regression using the NEX model to calculate the coefficients and intercepts for each season .

```
▶ x = TK_ssp585_NESM3_winter.values.reshape(-1, 1)
y = TK_winter.values.reshape(-1, 1)
reg = LinearRegression().fit(x, y)
print("r_square: ", reg.score(x,y))
print("coef: ", reg.coef_)
print("intercept: ", reg.intercept_)
```

```
⇨ r_square:  0.03251795094364984
coef:  [[-0.13680878]]
intercept:  [13.83076359]
```

```
[ ] x = TK_ssp585_NESM3_spring.values.reshape(-1, 1)
y = TK_spring.values.reshape(-1, 1)
reg = LinearRegression().fit(x, y)
print("r_square: ", reg.score(x,y))
print("coef: ", reg.coef_)
print("intercept: ", reg.intercept_)
```

```
r_square:  0.8850041217738248
coef:  [[1.40350478]]
intercept:  [9.44261699]
```

Winter & Spring

```
▶ x = TK_ssp585_NESM3_summer.values.reshape(-1, 1)
y = TK_summer.values.reshape(-1, 1)
reg = LinearRegression().fit(x, y)
print("r_square: ", reg.score(x,y))
print("coef: ", reg.coef_)
print("intercept: ", reg.intercept_)
```

```
⇨ r_square:  0.12647277625326792
coef:  [[0.29693856]]
intercept:  [41.24727687]
```

```
[ ] x = TK_ssp585_NESM3_fall.values.reshape(-1, 1)
y = TK_fall.values.reshape(-1, 1)
reg = LinearRegression().fit(x, y)
print("r_square: ", reg.score(x,y))
print("coef: ", reg.coef_)
print("intercept: ", reg.intercept_)
```

```
r_square:  0.9100590559523505
coef:  [[1.44500892]]
intercept:  [4.68581769]
```

Summer & Fall

Part 5. Conclusions & Recommendations

- Systematic difference identification
 - Potential practical usage
 - SSP 126 vs. SSP 585
-

Practical Usage of the Project

- The systematic difference estimation helps understand the gap between land surface temperature and near surface temperature.
- What's more important, it estimation gives us insight about how do we predict one from another, for example if the specific region we would like to obtain the land surface temperature (MODIS) has thousands row of missing value, which can occur in cloudy regions. we can estimate it using near surface air temperature from NEX.
- When choosing the NEX model, in most cases, two SSP scenarios return similar predicted value but when it comes to the outliers it is important to take the SSP scenarios into consideration because the prediction from SSP 585 is significantly higher than SSP 128

