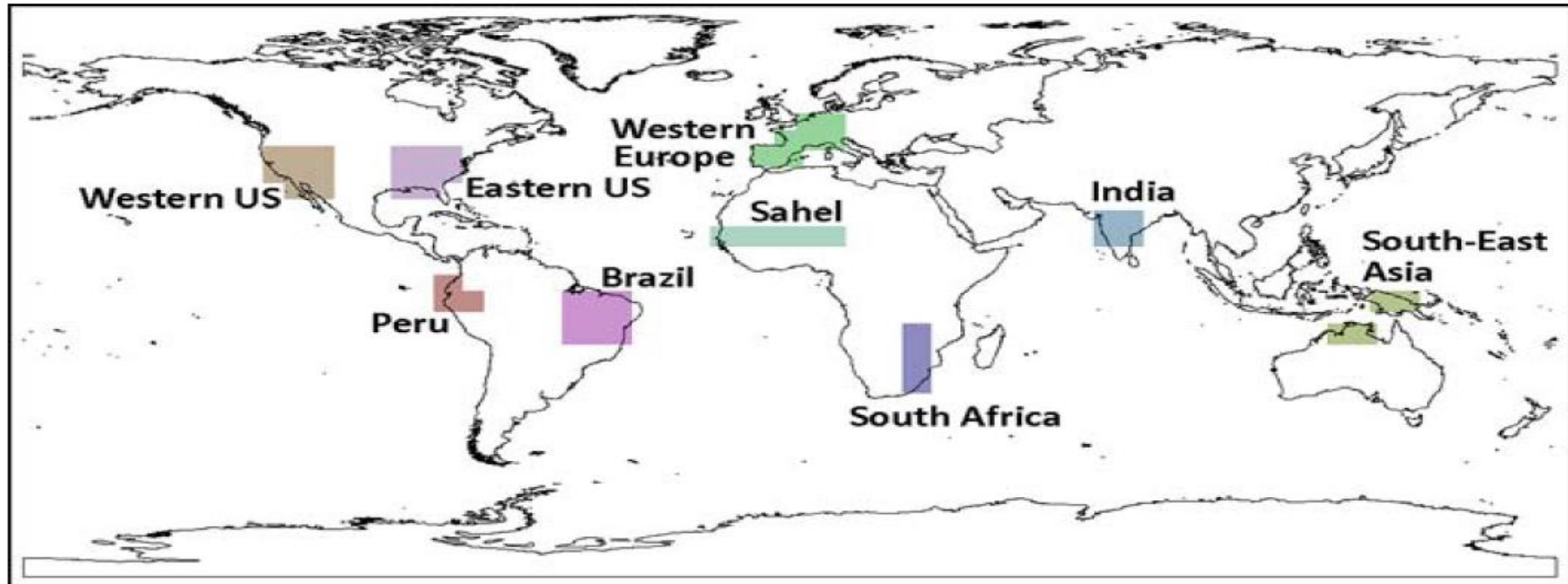


# Data Collection and Preprocessing

# Original Data – NCEP/NCAR Reanalysis 1 Dataset

- X – Predictor variables
  - Monthly means of temperature over ocean regions at  $2.5^\circ \times 2.5^\circ$  resolution
- y – Predictand variables
  - Air temperature over 9 land regions depicted below



# Original Data – NCEP/NCAR Reanalysis 1 Dataset

- Coordinates for y data

Region	Latitude coordinates	Longitude coordinates
Brazil	-20.0 - 0.0	300.0 - 320.0
Eastern USA	26.0 - 43.0	266.0 - 285.0
India	10.0 - 23.5	70.0 - 110.0
Peru	-8.0 - 3.5	279.0 - 287.0
South Africa	-33.0 - -11.0	20.0 – 35.0
Sahel	12.0 – 17.0	342.0 – 13.0
South East Asia	-18.0 - -2.0	127.0 – 147.0
Western Europe	36.0 – 50.0	352.0 – 12.0
Western USA	26.0 – 43.0	234.0 – 253.0

# Extract Corresponding Climate Model Data

- Website: <https://pcmdi.llnl.gov/search/cmip5/>
- Dataset details
  - Project: CMIP5
  - Experiment: historical
  - Experiment Family: All
  - Time Frequency: mon
  - Realm: X – ocean, y - land
  - Variable: X – tos, y - tas

# Dataset Details

- $X \in R^{24612 \times 5881}$  - Model sea surface temperature at resolution 2.5 x 2.5
- $y \in R^{24612 \times 9}$  - Surface air temperature variables over 9 locations
- $\text{latlon} \in R^{144 \times 72}$  - Locations of X input, 0 indicates no data
- Climate models

Row numbers	Model
1-1872	CMCC-CESM
1873-3744	CMCC-CM
3745-5616	CMCC-CMS
5617-7488	FGOALS-g2
7489-9348	FGOALS-s2-r2i1p1
9349-11208	FGOALS-s2-r3i1p1
11209-12960	HadGEM2-AO
12961-14832	INMCM4
14833-16788	MIROC5-r1i1p1
16789-18744	MIROC5-r2i1p1
18745-20700	MIROC5-r3i1p1
20701-22656	MIROC5-r4i1p1
22657-24612	MIROC5-r5i1p1

# Data Preprocessing

- For each model, each location
  - Compute means and standard deviation for each month separately
  - Zscore monthly data