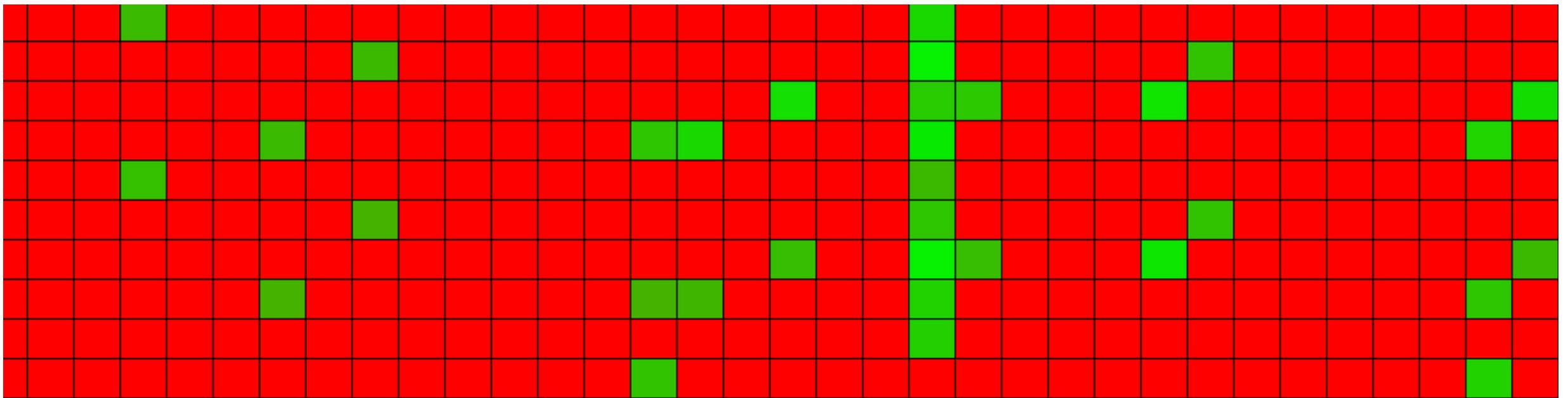


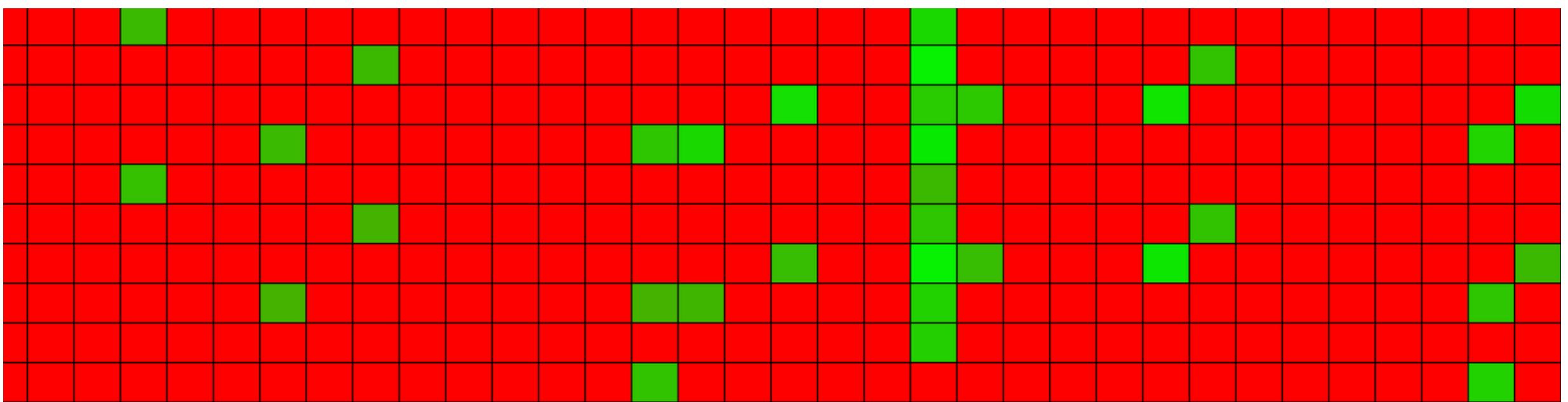
# Improving sub-seasonal drought forecasting with machine learning and climate indices

IHE Delft, 08/12/2022





you can find  
the slides  
here!



# Today's Agenda

this presentation will go through the following stages:

01

Intro

02

Context

03

Framework

04

Next steps

# Intro

- 01 What is drought
- 02 ML for Drought
- 03 The gap

# Intro

- 01 What is drought
- 02 ML for Drought
- 03 The gap

## Meteorological Drought

a period of time in which a region experiences below-normal precipitation

Reduced soil moisture, Reduced stream flow, Crop damage

## Water shortage

The onset, extent and duration of drought are difficult to define

different stakeholders have varying degrees of tolerance and resilience to these events

(Slette et al., 2019)

Being able to forecast them is crucial

# Intro

exploitation of **statistic** and **dynamic techniques** for droughts forecasting

has been and is widely studied

**sub-seasonal forecasting**

- 01 What is drought
- 02 ML for Drought
- 03 The gap

Earth observation data

Artificial Intelligence  
Hardware (GPU,TPU)

McGovern et al. (2017)

AI-based  
prediction  
models

Learn from past data  
Integrate physical understanding into the models  
Discover additional knowledge from the data  
Handle large amounts of input variables

# Intro

- 01 What is drought
- 02 ML for Drought
- 03 The gap

Why to focus on  
sub-seasonal  
lead times?

## Informative predictors

seasonal:

climate indices and large scale teleconnection patterns

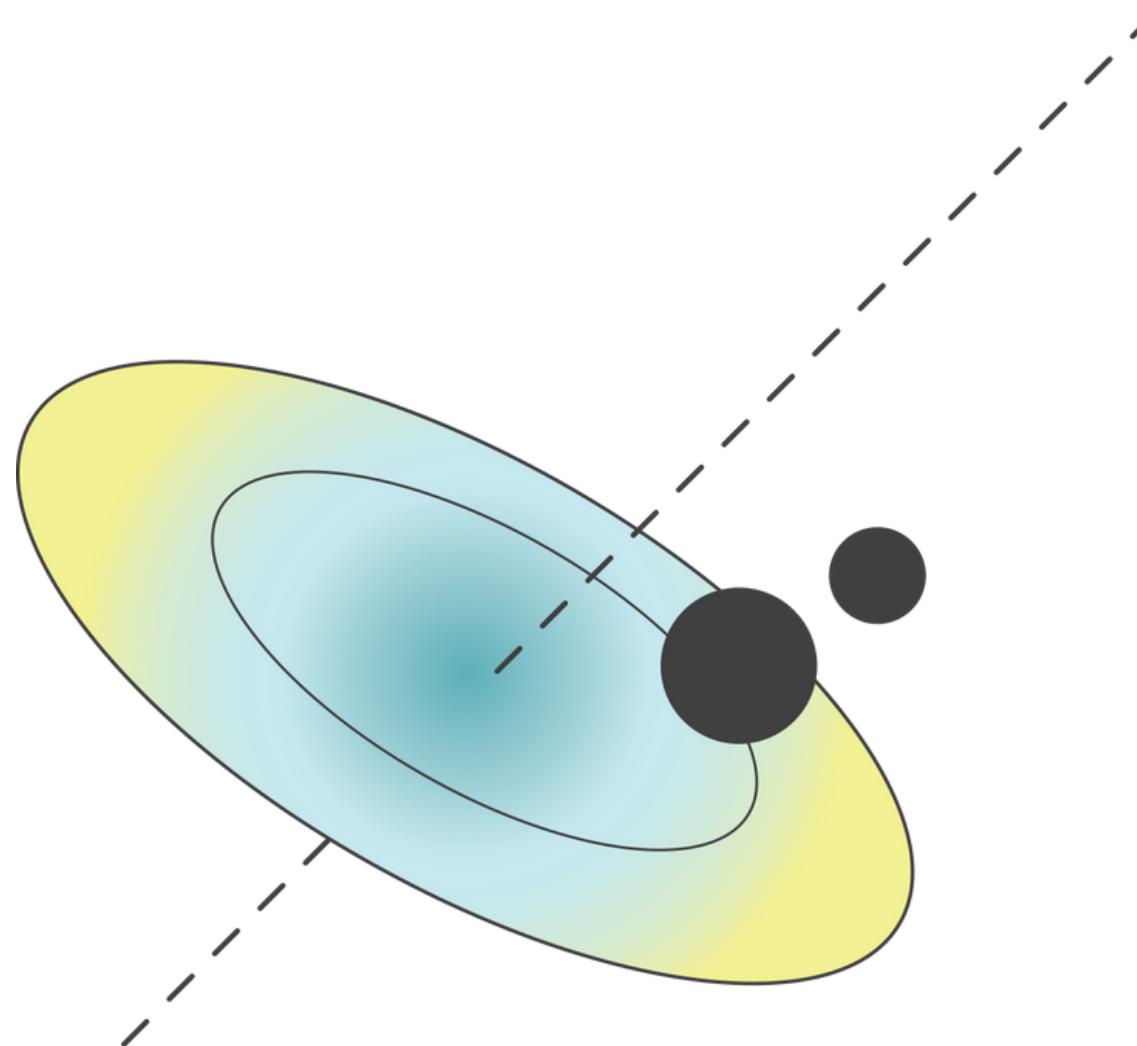
short-medium term:

local variable (precipitation, temperature)

sub-seasonal?

- short enough that the atmosphere still has memory of its initial conditions
- long enough to allow atmospheric circulation to affect the evolution of weather conditions

# Context



- 01 What (our goal)
- 02 Where (study area)
- 03 How (the framework)

# Context

Machine Learning model for  
sub-seasonal precipitation  
forecasting

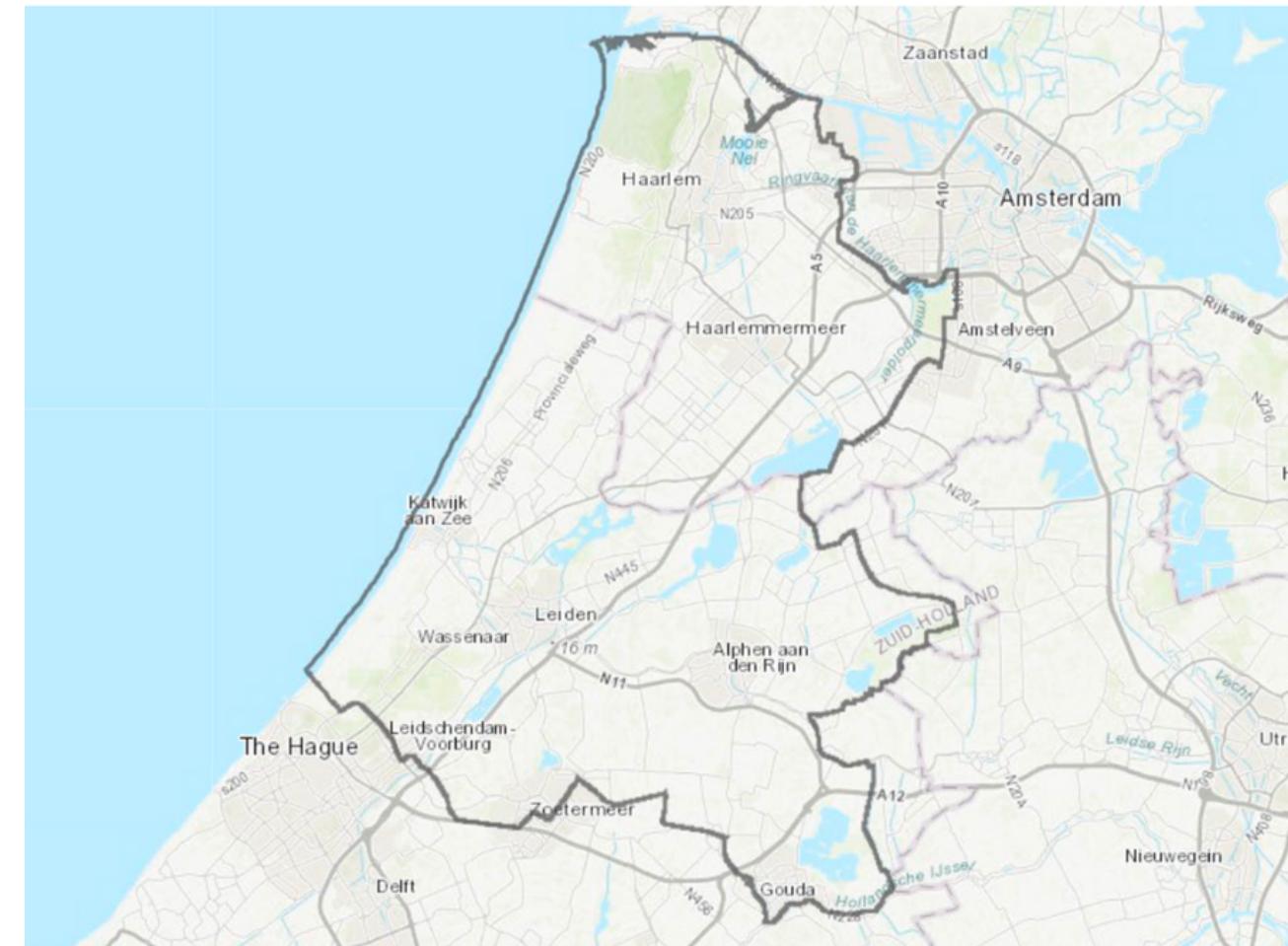
precipitation  
forecasting



drought  
forecasting

- 01 What (our goal)
- 02 Where (study area)
- 03 How (the framework)

# Context



- 01 What (our goal)
- 02 Where (study area)
- 03 How (the framework)

## Rijnland

small sub-catchment of 1000 km<sup>2</sup> at the very end of the Rhine delta in the Netherlands

water board of Rijnland is able to forecast drought at **bi-weekly** lead times. The goal is to extend it to **a month**

# Context

- 01 What (our goal)
- 02 Where (study area)
- 03 How (the framework)

**Nino Index Phase Analysis  
(NIPA)**

Zimmerman et al. (2016)



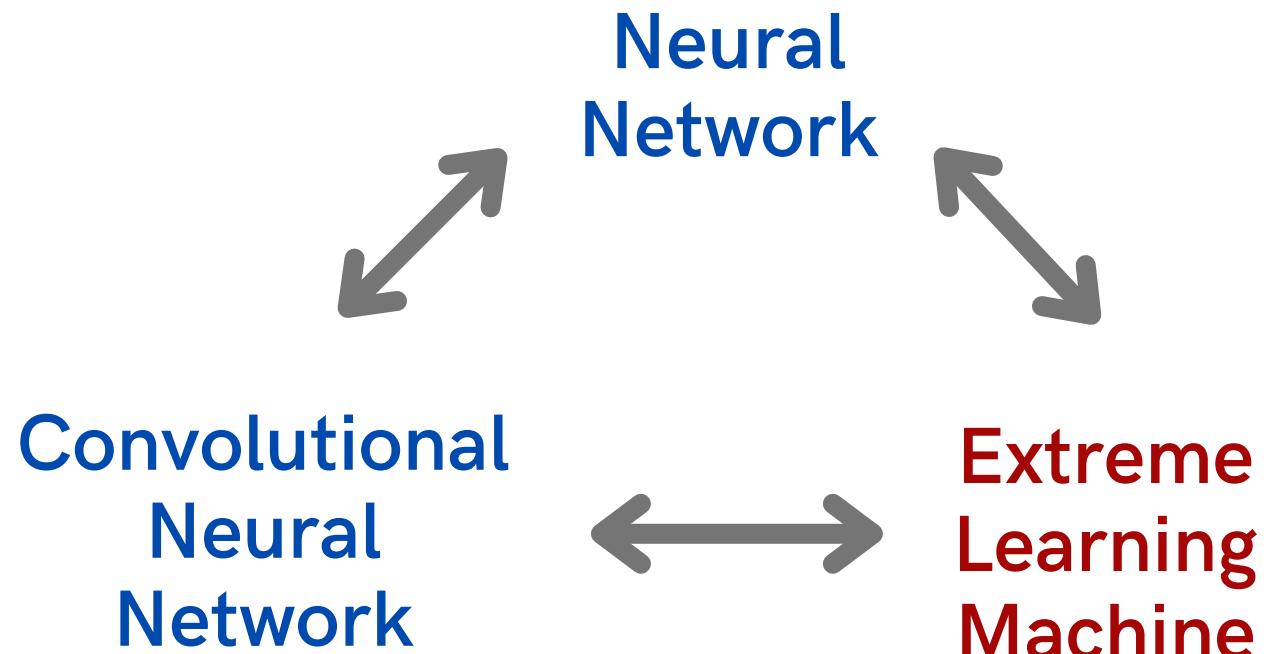
Giuliani et al. (2019)



Our readaptation



**Neural Networks**



# Framework

- 01 NIPA
- 02 ELM

# Framework

NIPA is a framework that searches for links between **Global** and **Local variables** exploiting the phases of teleconnection patterns materialized by **climate indices**

- 01 NIPA
- 02 ELM

**climate indices**

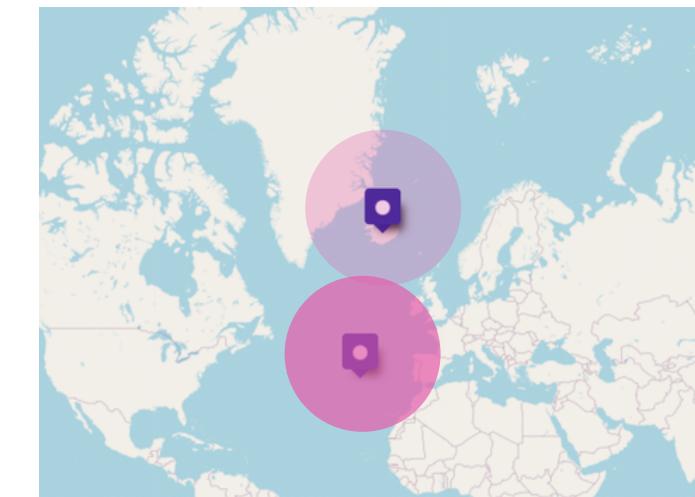
El Niño Southern Oscillation (ENSO)  
North Atlantic Oscillation (NAO)  
SCAndinavian oscillation (SCA)  
East Atlantic oscillation (EA)

# Framework

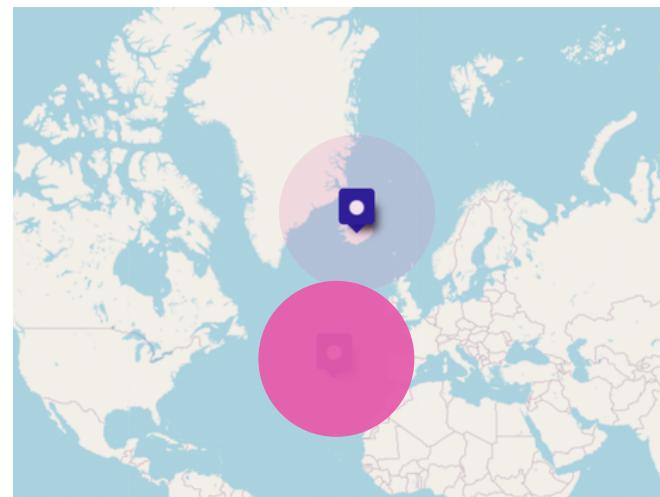
- above/below-normal temperatures in eastern United States and northern Europe
- above/below-normal temperatures in Greenland and southern Europe
- above/below-normal precipitation over northern Europe and Scandinavia
- above/below-normal precipitation over southern and central Europe

- 01 NIPA
- 02 ELM

climate indices  
North Atlantic Oscillation (NAO)



Phase Neg



Phase Pos

# Framework

- 01 NIPA

- 02 ELM

Input

## DATA

- Local precipitation (monthly timeseries) - cumulative
- Global variable (monthly timeseries) - SLP,SST,Z500 - mean
- Climate Index (monthly timeseries) - ENSO, NAO,SCA,EA

Data extraction

Phase segmentation

Correlation

PCA

output

## SETTING PARAMETERS

- Month (of local precipitation)
- Aggregation level (of pre-month global data)

ERA5

# Framework

- 01 NIPA
- 02 ELM

## SETTING PARAMETERS

- Month **(of local precipitation)**
- Aggregation level **(of pre month global data)**

### Example:

- Month **1** local precipitation of **January** and the global variable of **December**
- Month **1** local precipitation of **January** and the global variable of **November + December**

Input

Data extraction

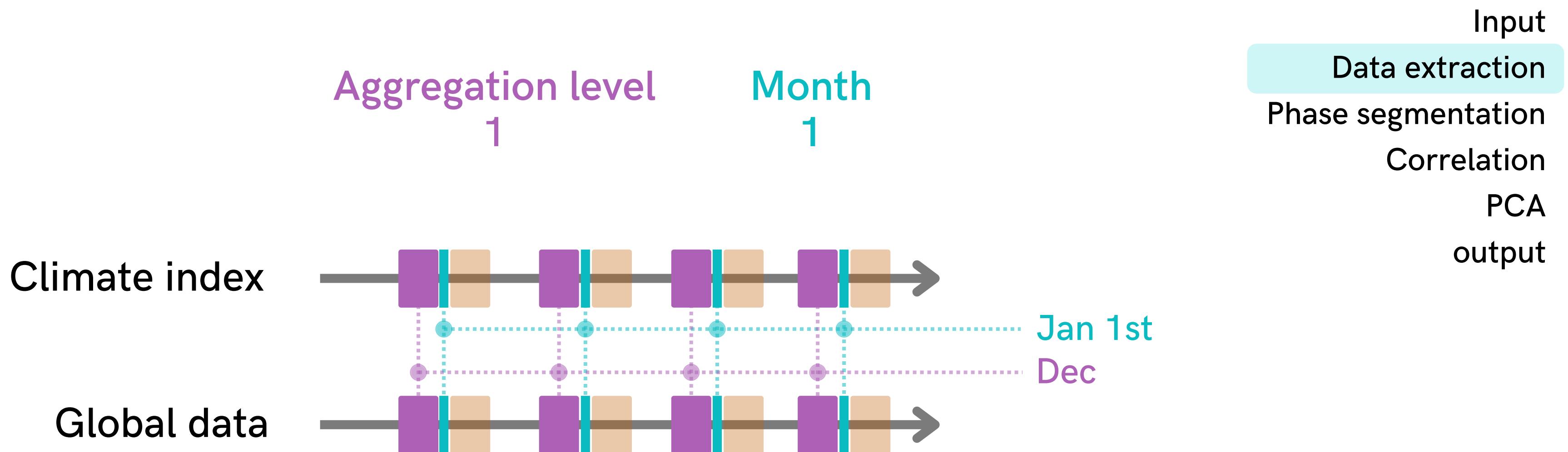
Phase segmentation

Correlation

PCA

output

# Framework

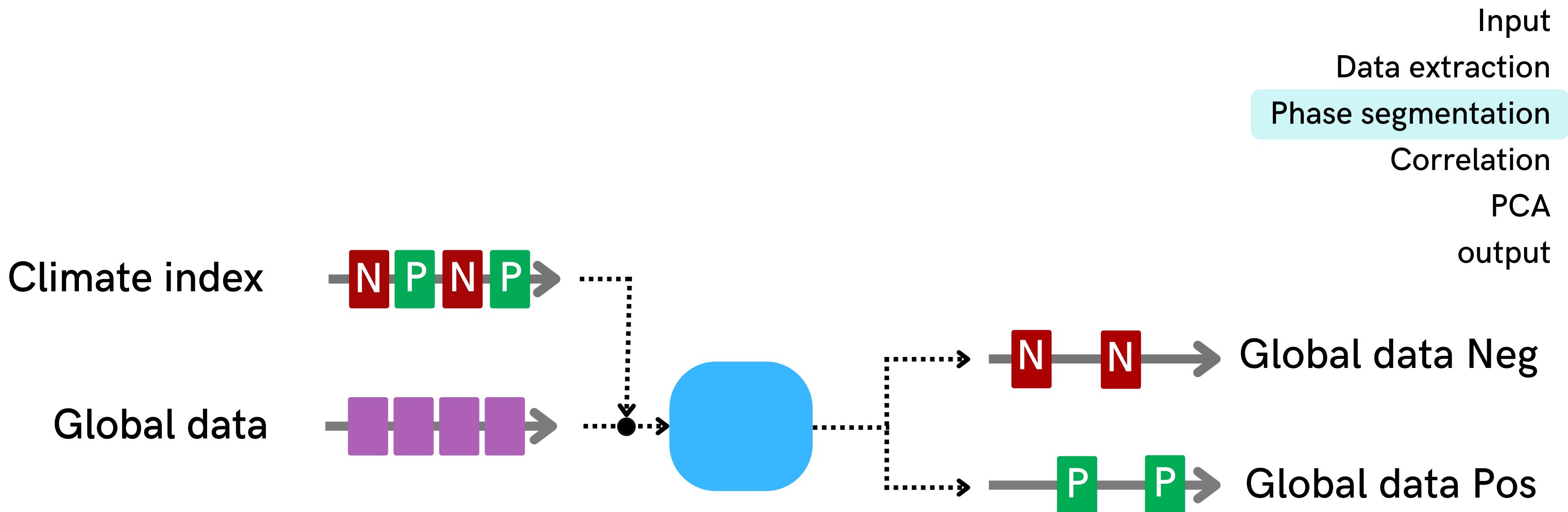


● 01 NIPA  
● 02 ELM

Input  
Data extraction  
Phase segmentation  
Correlation  
PCA  
output

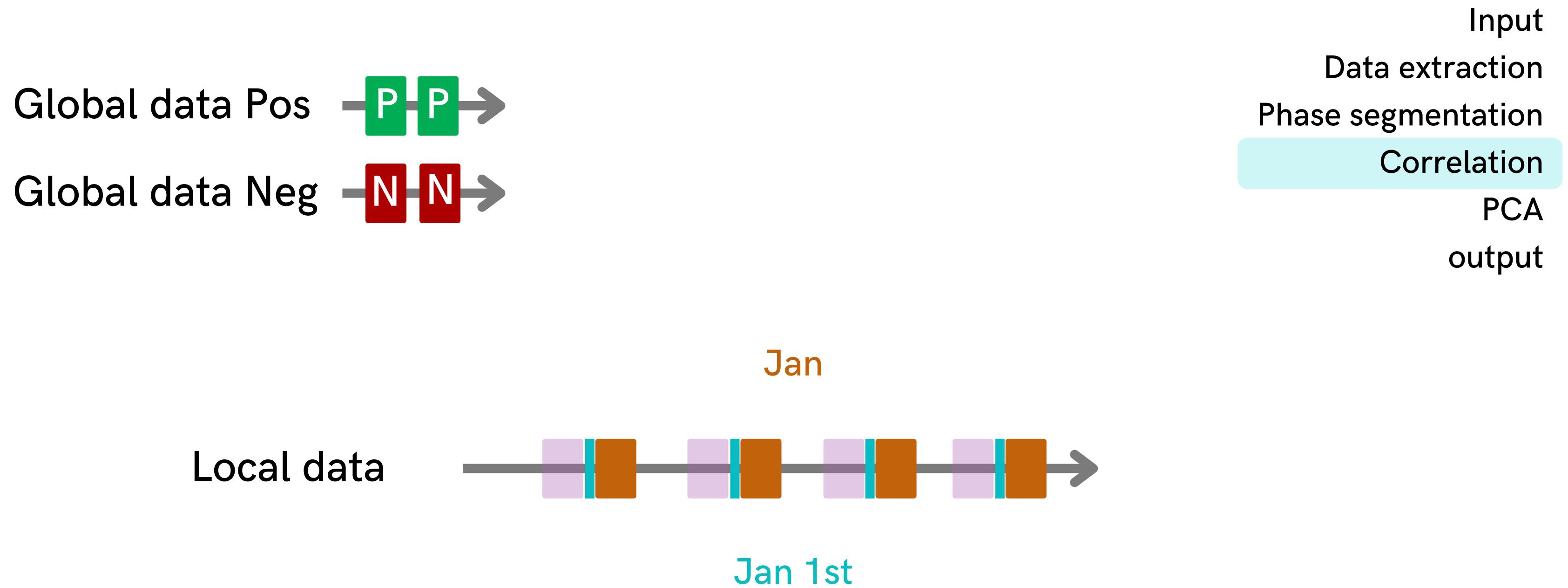
NOTE: this is an year-based operation. NIPA will extract the data for the December of each year

# Framework

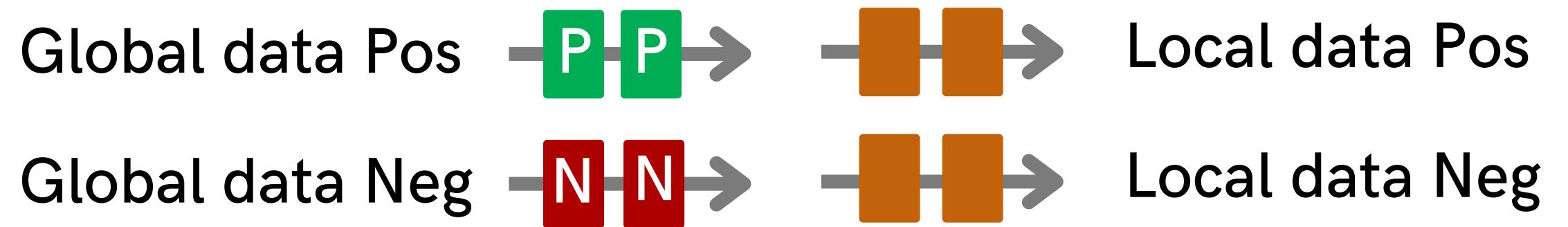


# Framework

- 01 NIPA
  - 02 ELM
- 



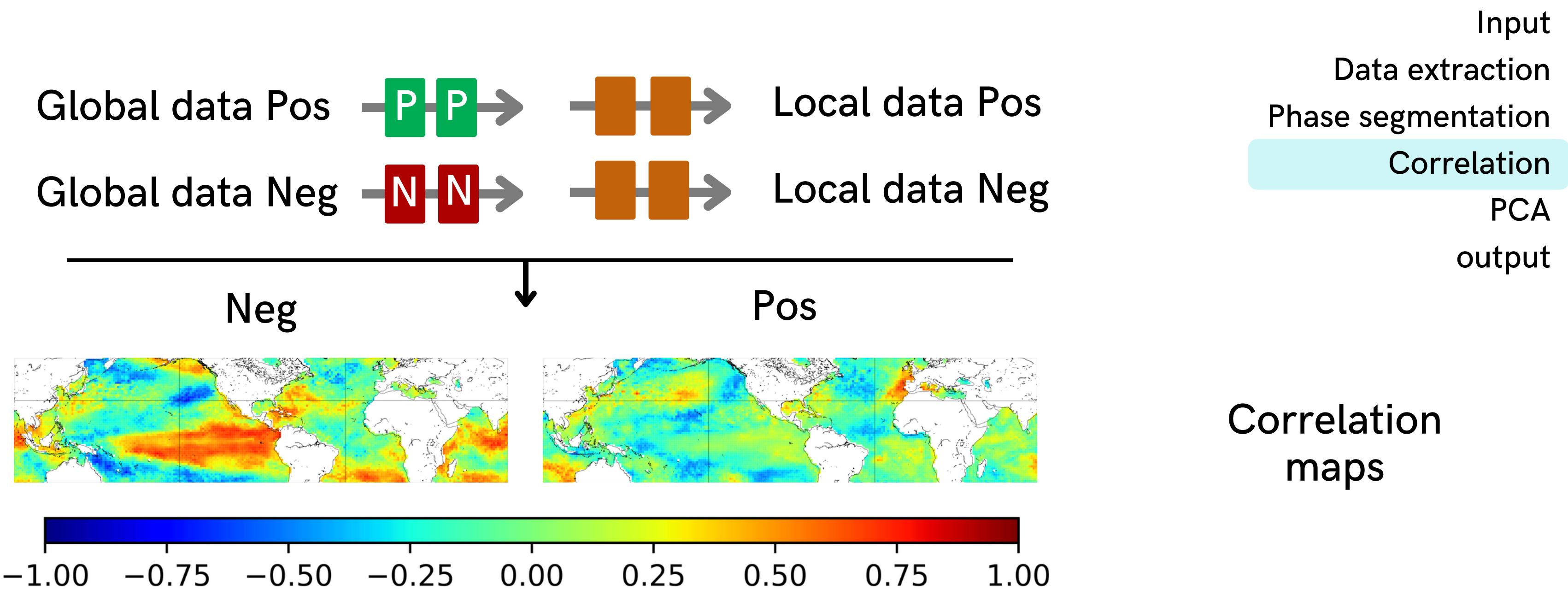
# Framework



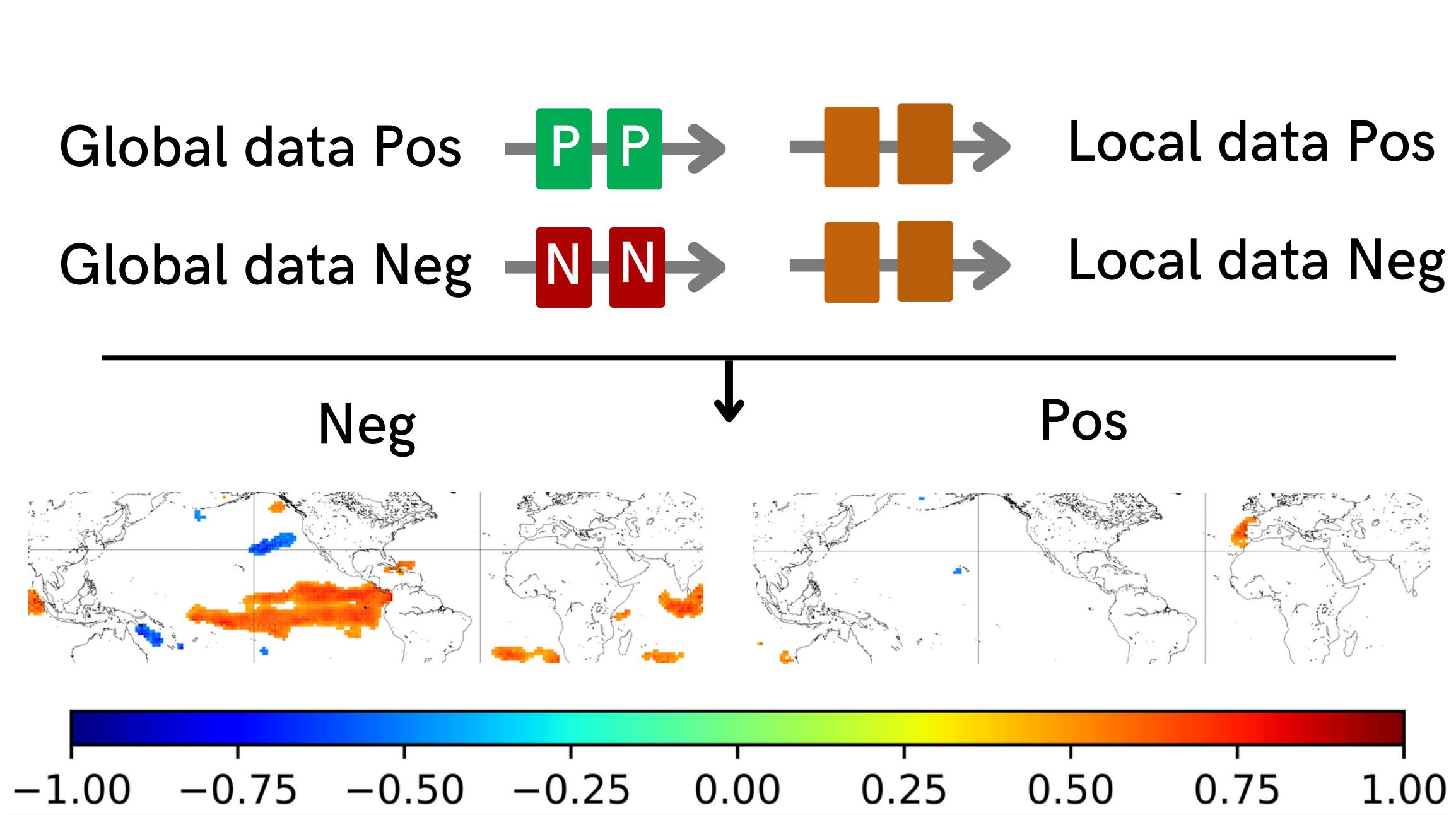
- 01 NIPA
- 02 ELM

Input  
Data extraction  
Phase segmentation  
Correlation  
PCA  
output

# Framework



# Framework

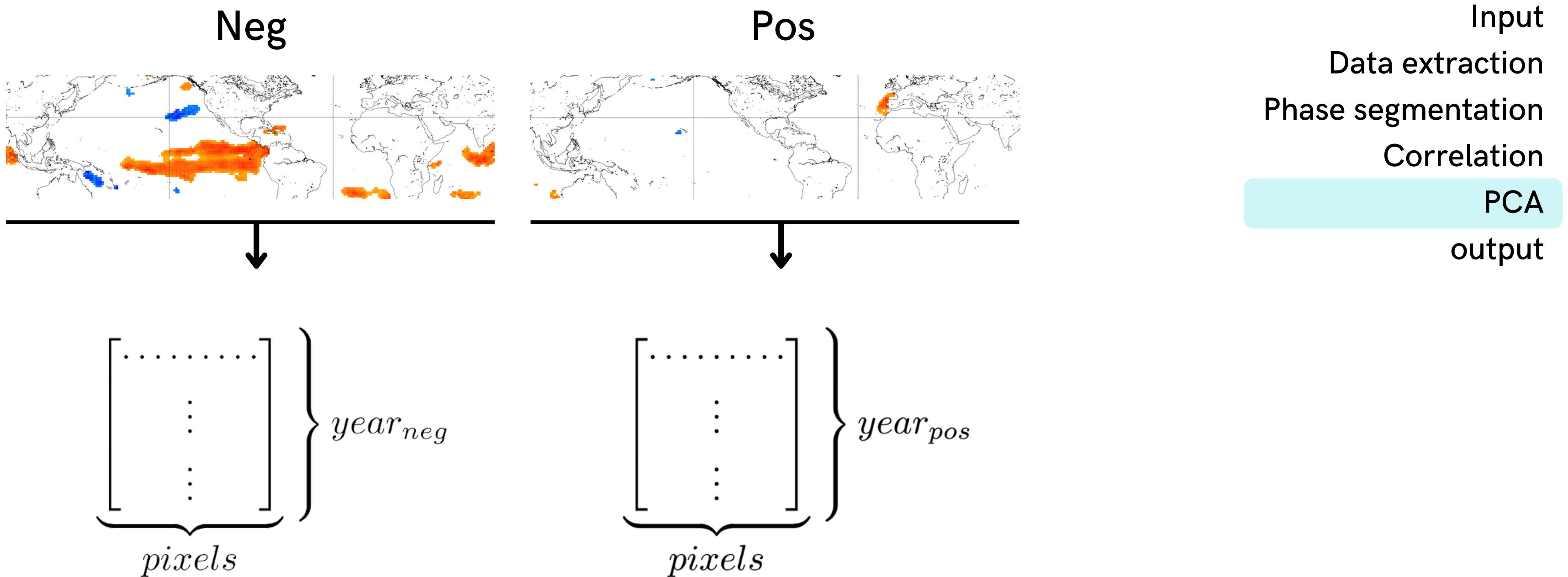


- 01 NIPA
- 02 ELM

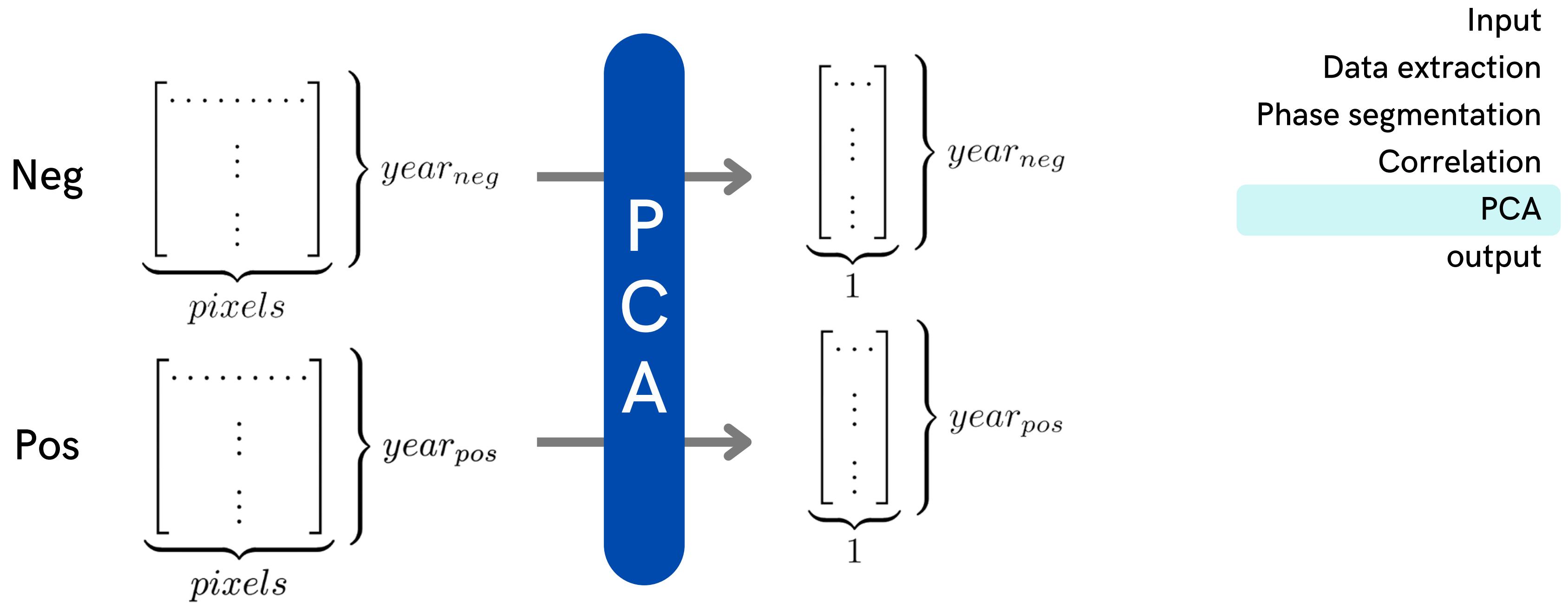
Input  
Data extraction  
Phase segmentation  
Correlation  
PCA  
output

**95% of significance**  
+  
**minimum correlation threshold 0.6**  
+  
**3x3 contiguous area check**

# Framework



# Framework



# Framework

PC1	phase_label
PC1 1979	1
PC1 1980	2
...	...
...	...
PC1 2021	2

- 01 NIPA
- 02 ELM

Input  
Data extraction  
Phase segmentation  
Correlation  
PCA  
output

Dataset for  
1 month

# Framework

This procedure can be applied

- for each **Month**
- for each combination of:
  - **Local Precipitation**
  - **Global Variable (SST/SLP/Z500)**
- for each **aggregation level** of SST/SLP/Z500 (1/2/3 month)
- for each **climate signal** (ENSO/NAO/SCA/EA)

After all the running of NIPA: **432 datasets**.

By applying the three filtering conditions : **101 datasets**

● 01 NIPA

● 02 ELM

Input

Data extraction

Phase segmentation

Correlation

PCA

output

# Framework

● 01 NIPA

● 02 ELM

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500
SCA_N	0	0	0	0	0	-0,67	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0
ENSO-me_i_N	0	0	0	0	0	0	0	0	0	-0,82	0	0
NAO_N	0	0	0	0	0	-0,67	0	0	0	0	0	0
SCA_P	0	0	0	0	-0,69	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	-0,64	0	0	0	0	0
ENSO-me_i_P	0	0	0	0	0	0	0	-0,87	0,66	0	0	0
NAO_P	0	0	0	0	-0,64	0	0	0,73	0	0	0	-0,66
SCA_N	0	0	0	0	0	0	0	0,75	0	0	0	0
EA_N	0	0	0	0	0	0	0	0,84	0	0	0	0
ENSO-me_i_N	0	0	0	0	0	0	0	0,71	0,7	0	0	0
NAO_N	0	0	0	0	0	0	0	0,81	0	0	0	0
SCA_P	0	0	0	0	0	0	0	-0,66	0	0	0	0
EA_P	0	0	0	0	0	0	-0,64	0	0	0	0	0
ENSO-me_i_P	0	0	0	0	0	0	0	-0,87	0,66	0	0	0
NAO_P	0	0	0	0	0	0	0	0,73	0	0	0	0,74
SCA_N	0	0	0	0	0	0	0	0,72	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0,73
ENSO-me_i_N	0	0	0	0	0	0	0	0	0,78	0	0	0
NAO_N	0	0	0	0	0	0	0	0,84	0	0	0	-0,65
SCA_P	0	0	0	0	0	0	0	0,65	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0,66
ENSO-me_i_P	0	0	0	0	0	0	0	0	0,76	0	0	0
NAO_P	0	0	0	0	0	0	0	-0,71	0	0	0	0,74
SCA_N	0	0	0	0	-0,79	0	0	0	0,66	0	0	0
EA_N	0	0	0	0	0	0	0	0,7	0	0	0	-0,77
ENSO-me_i_N	0	0	0	0	-0,7	0	0	-0,68	0,72	0	0	0
NAO_N	0	0	0	0	0	0	0	0	-0,67	0	0	0
SCA_P	0	0	0	0	0,68	0	0	0	0	-0,76	0	0
EA_P	0	0	0	0	0	0	0	-0,65	0	0	0	0,63
ENSO-me_i_P	0	0	0	0	-0,76	0	0	-0,69	-0,63	0	0	-0,71
NAO_P	0	0	0	0	0	0	0	0	-0,73	0	0	0

# Framework

- 01 NIPA
  - 02 ELM
- 

Local Data

The Local Data have been obtained starting from the ERA5 dataset (as the global data and the target).

The data consists of timeseries of 11 different variables referred only to the Rijnland grid cell (re-gridded on ECMWF):

Cumulative precipitation (tp)  
2m temperature (t2m)  
Total Cloud Cover (TCC)  
Mean Evaporation Rate (MER)  
Mean Surface Sensible Heat Flux (MSSHF)

Snow Depth (SD)  
U-component of wind (UW)  
V-component of wind (VW)  
Relative Humidity (RH)  
Specific Humidity (SH)  
Total Column Water Vapour (TCWV)

Felsche  
et.al.

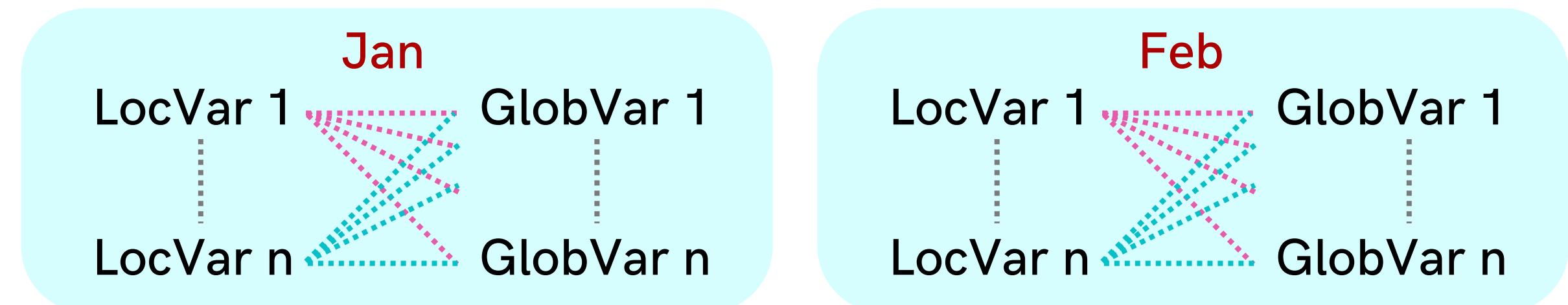
# Framework

Creation of **monthly-based** datasets with all the possible combinations between **local variables** and **global variables** (if present).

Combinations constraints:

- Maximum **2 global variables** considered for a single dataset
- Maximum **4 variables** for a single dataset (global + local)

in total 13.541 combinations (datasets) have been created



- 01 NIPA
- 02 ELM

Local Data  
Local/Global combinations  
Model Selection  
Final Results  
Comparison

# Framework

Application of a *Leave One Out* (LOO) model selection procedure to:

- Select the **most informative** mix of **global** and **local variables** *for each month*.
- Select the **best number of neurons** in the hidden layer of the ELM models *for each month*
- Select the **best activation function** for the neurons of the hidden layers of the ELM models *for each month*

- 01 NIPA
- 02 ELM

Local Data

Local/Global combinations

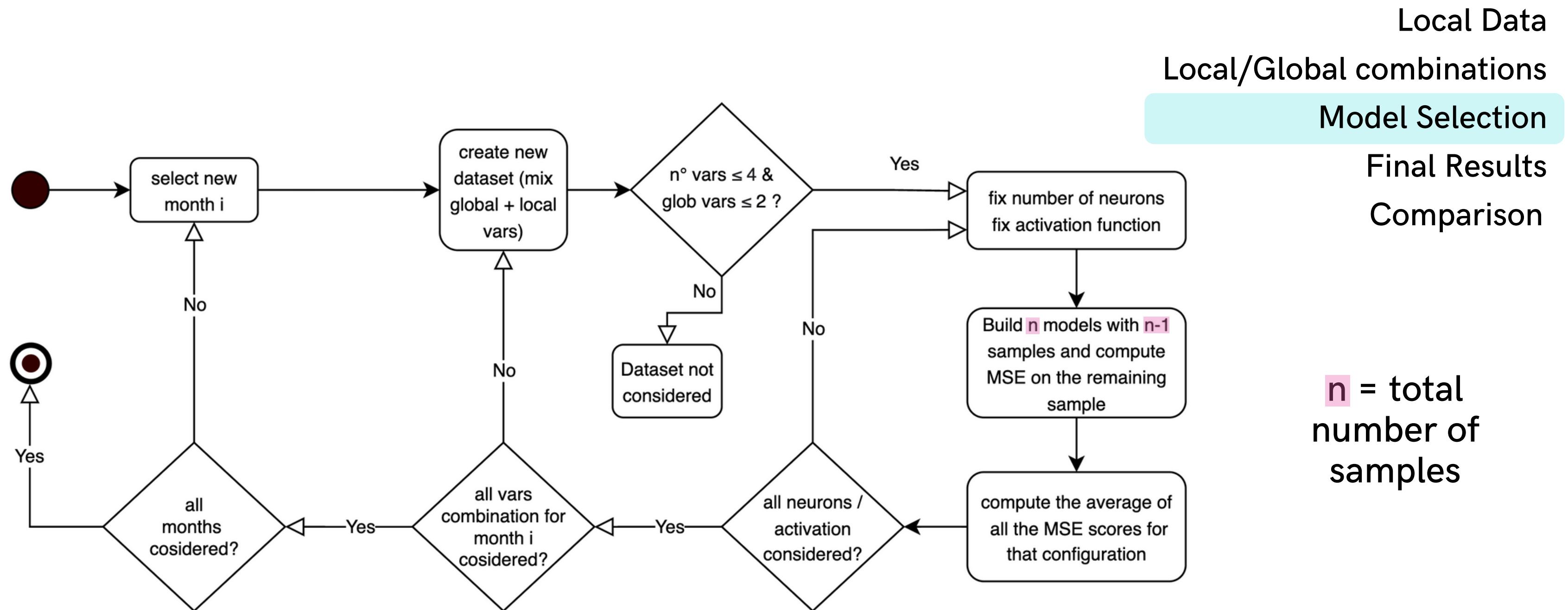
Model Selection

Final Results

Comparison

# Framework

- 01 NIPA
- 02 ELM



# Framework

	1	2	3	4	5	6	7	8	9	10	11	12
t2m-TCC-TCWV-UW	1136.77	1096.17	993.6	1033.96	1235.78	1083.2	1769.69	2169.56	1999.84	1509.22	855.37	1030.84
t2m-TCC-TCWV-VW	792.45	1157.07	896.2	932.24	1023.3	1176.6	2114.29	1720.88	2205.18	1796.66	849.73	755.65
t2m-TCC-tp-UW	1038.49	1088.63	977.3	1015.95	1216.38	1247.02	1853.87	1955.14	1978.42	1405.76	750.29	1007.6
t2m-TCC-tp-VW	1231.12	1356.22	1003.62	921.48	1131.77	1183.61	1806.75	1754.28	2209.74	1390.09	879.6	1209.71
t2m-TCC-UW-VW	985.11	1139.13	968.54	1126.59	1099.41	1203.25	1810.71	1855.56	1682.82	1437.11	822.99	1084.57
t2m-TCWV-tp-UW	1001.08	1115.88	1203.79	908.28	1030.19	1154.62	1806.95	1918.36	1996.25	1334.13	813.21	957.55
t2m-TCWV-tp-VW	1101.01	1205.01	1091.0	1087.14	1181.95	1299.88	2272.37	2074.43	1757.31	1424.22	769.86	903.9
t2m-TCWV-UW-VW	1071.39	1158.17	994.16	941.98	1112.11	1089.73	2070.87	2150.51	1733.09	1372.57	935.32	831.81
t2m-tp-UW-VW	683.74	1051.2	793.15	899.56	1135.57	1211.52	1964.32	2087.7	1802.9	1565.64	742.54	991.9
TCC-TCWV-tp-UW	895.82	1156.45	951.0	1025.64	1066.09	1229.83	1928.64	2001.44	1568.77	1634.16	837.77	767.73
TCC-TCWV-tp-VW	1070.19	895.15	1078.36	1066.25	1198.34	1107.59	1555.76	2036.93	2074.02	1672.47	881.7	941.32
TCC-TCWV-UW-VW	1202.89	1343.99	1037.63	1105.42	945.94	1011.13	1834.21	1986.2	1866.81	1212.18	909.2	923.24
TCC-tp-UW-VW	971.23	993.72	894.99	1150.58	1200.45	1054.45	1616.44	1567.07	1570.86	1444.48	914.89	1164.83
TCWV-tp-UW-VW	1021.35	1183.97	941.41	959.78	939.39	1237.65	1522.18	2531.57	1458.46	1493.15	897.63	1143.04
SCA_Z500-1_tp-2_dataset.csv	500.97											
MER-SCA_Z500-1_tp-2_dataset.csv	592.11											
MSSHF-SCA_Z500-1_tp-2_dataset.csv	743.77											
RH-SCA_Z500-1_tp-2_dataset.csv	1051.21											
SD-SCA_Z500-1_tp-2_dataset.csv	590.71											
SH-SCA_Z500-1_tp-2_dataset.csv	593.59											
t2m-SCA_Z500-1_tp-2_dataset.csv	571.29											
TCC-SCA_Z500-1_tp-2_dataset.csv	545.17											

- 01 NIPA
- 02 ELM

Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

one table for each  
combination of  
neurons/activation

functions (reporting the  
LOO validation MSE)

22 tables

# Framework

- 01 NIPA
- 02 ELM

Select one setting for each month having the best LOO  
Validation Error

```
1: (667.93, 9, 'relu', 'SH-t2m-UW-VW'),  
2: (425.97, 5, 'relu', 'MSSHF-TCC-VW-SCA_Z500-1_tp-2_dataset.csv'),  
3: (428.32, 8, 'relu', 'RH-UW-NAO_Z500-1_tp-3_dataset.csv'),  
4: (384.6, 12, 'sigm', 'tp-SCA_MSLP-3_tp-4_dataset.csv'),  
5: (672.03, 4, 'sigm', 'MER-RH-SH-tp'),  
6: (193.40, 5, 'sigm', 'SD-TCC-NAO_MSLP-1_tp-6_dataset.csv-EA_MSLP-2_tp-6_dataset.csv'),  
7: (676.80, 4, 'sigm', 't2m-TCC-ENSO-meい_MSLP-1_tp-7_dataset.csv'),  
8: (380.20, 9, 'relu', 'SH-tp-SCA_MSLP-2_tp-8_dataset.csv-ENSO-meい_Z500-2_tp-8_dataset.csv'),  
9: (698.42, 10, 'sigm', 'MER-SD-SCA_MSLP-3_tp-9_dataset.csv'),  
10: (418.51, 12, 'sigm', 'SH-ENSO-meい_SST-1_tp-10_dataset.csv'),  
11: (360.01, 12, 'sigm', 'MSSHF-EA_MSLP-3_tp-11_dataset.csv'),  
12: (266.99, 10, 'sigm', 'MSSHF-VW-NAO_MSLP-1_tp-12_dataset.csv-EA_MSLP-2_tp-12_dataset.csv')
```

Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

# Framework

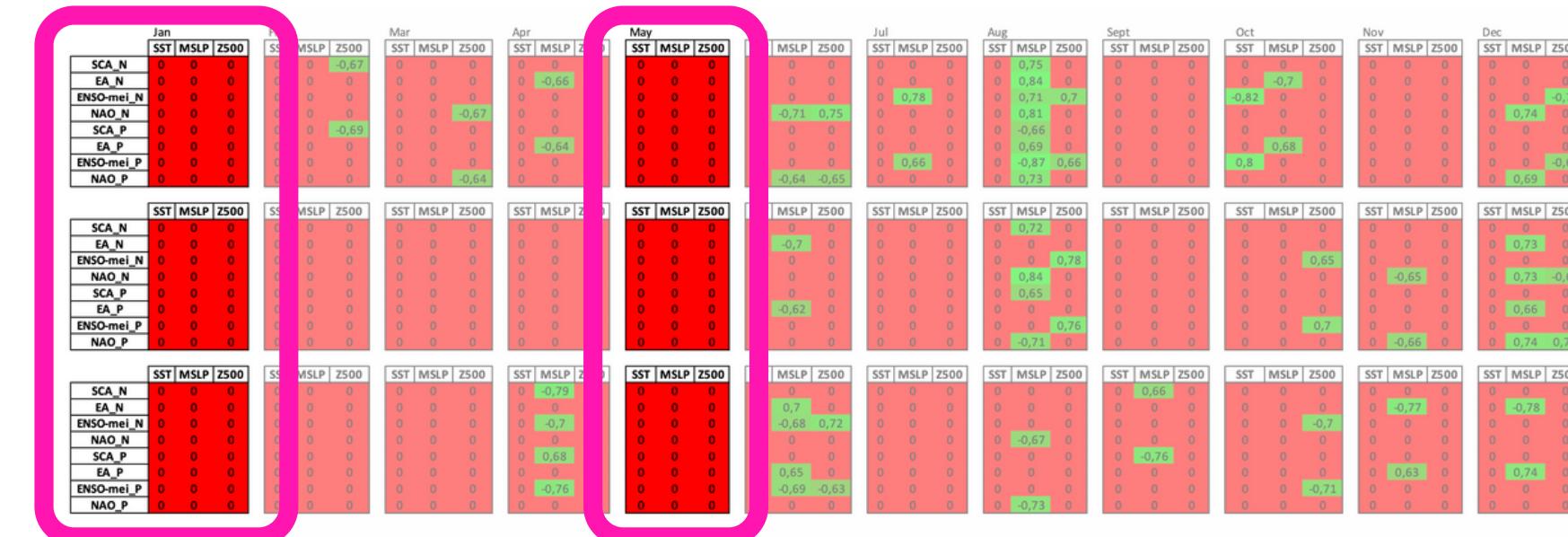
• 01 NIPA

• 02 ELM

Select one setting for each month having the best LOO  
Validation Error

- 1: (667.93, 9, 'relu'),
- 2: (425.97, 5, 'relu'),
- 3: (428.32, 8, 'relu'),
- 4: (384.6, 12, 'sigm'),
- 5: (672.03, 4, 'sigm'),
- 6: (193.40, 5, 'sigm'),
- 7: (676.80, 4, 'sigm'),
- 8: (380.20, 9, 'relu'),
- 9: (698.42, 10, 'sigm'),
- 10: (418.51, 12, 'sigm'),
- 11: (360.01, 12, 'sigm'),
- 12: (266.99, 10, 'sigm')

months with no NIPA output (global climate context not considered to build the ELM). ELM must rely only on local data to make predictions



Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

# Framework

- 01 NIPA
- 02 ELM

Select one setting for each month having the best LOO Validation Error

```

1: (667.93, 9, 'relu'),
2: (425.97, 5, 'relu'),
3: (428.32, 8, 'relu'),
4: (384.6, 12, 'sigm'),
5: (672.03, 4, 'sigm'),
6: (193.40, 5, 'sigm'),
7: (676.80, 4, 'sigm'),
8: (380.20, 9, 'relu'),
9: (698.42, 10, 'sigm'),
10: (418.51, 12, 'sigm'),
11: (360.01, 12, 'sigm'),
12: (266.99, 10, 'sigm')

```

months with just 1 NIPA output (out of 12 possible). ELM has only one option to consider in the global climate context. Probably the relevant climate signals for these months do not fall into the set we have considered



Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

# Framework

Build a model for each moth based on the best setting



- 01 NIPA
- 02 ELM

Local Data  
Local/Global combinations  
Model Selection  
Final Results  
Comparison

# Framework

- 01 NIPA
  - 02 ELM
- 

## Giuliani et. al.

- 12 models (1 for each month)
- n° neurons: 10
- act. function: sigmoid

## Our work

- 12 models (1 for each month)
- n° neurons: range (4,12)
- act. function: (sigmoid,relu)

↓  
Pearson  
prediction vs target:  
**0.81**

↓  
Pearson  
prediction vs target:  
**0.81/0.82**

Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

# Framework

- The comparison is based on Pearson because is the same metric used in the paper (no MSE, RMSE, etc. is provided)
- The mix of local/global vars seems to be effective in slightly improving the predictive skills
- Better results could be obtained by considering more climate signals to
  - produce outputs for Jan and May
  - produce more than 1 output for Jul and Sep
- Only LOO validation without testing to be consistent with Giuliani et. al. (low number of samples )

● 01 NIPA

● 02 Neural Network

Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

# Next steps

- 01 Neural Network
- 02 Conv. Neural Net.

# Next steps

- 01 Neural Network
  - 02 Conv. Neural Net.
- 

Because of the **lack of samples** and the absence of a **proper testing procedure**, we plan to:

- build a **single neural network** for the whole period
- **compare it month by month** with the monthly-based ELMs

The reason why is to **check** if **neglecting NIPA** (which checks for dependencies between variables through phases of climate indices) could be **compensated by the presence of more training samples** able to make the neural network learn (part of) the underlying patterns by itself

# Next steps

- 01 Neural Network
  - 02 Conv. Neural Net.
- 

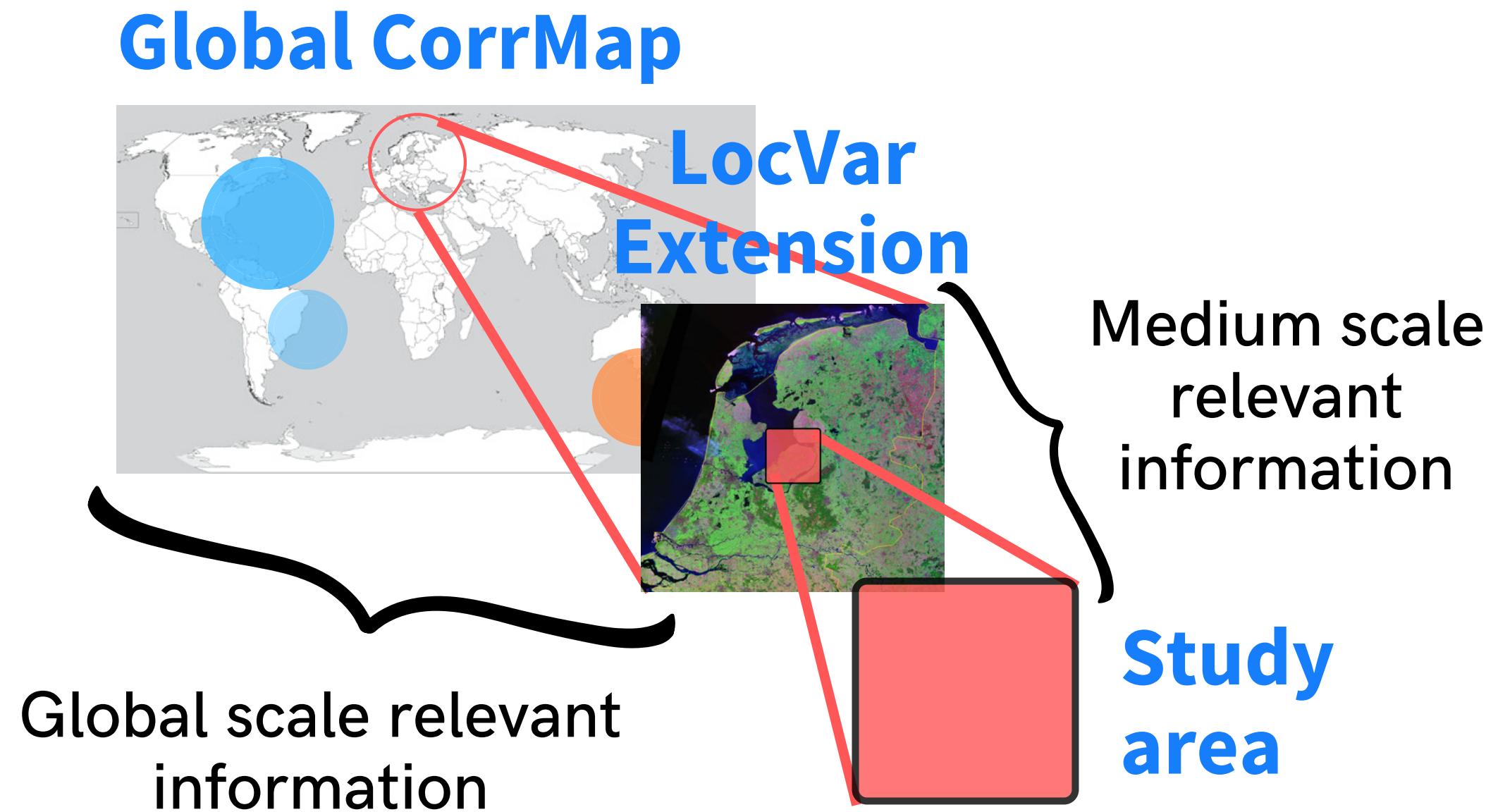
Because of the **lack of samples** and the absence of a **proper testing procedure**, we plan to:

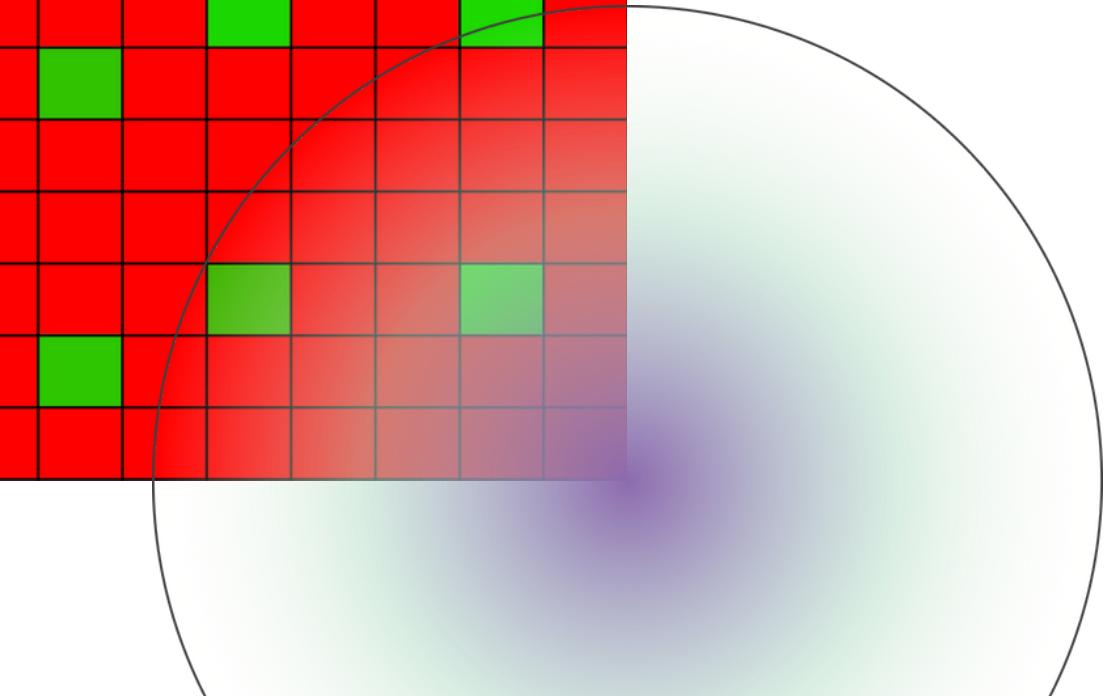
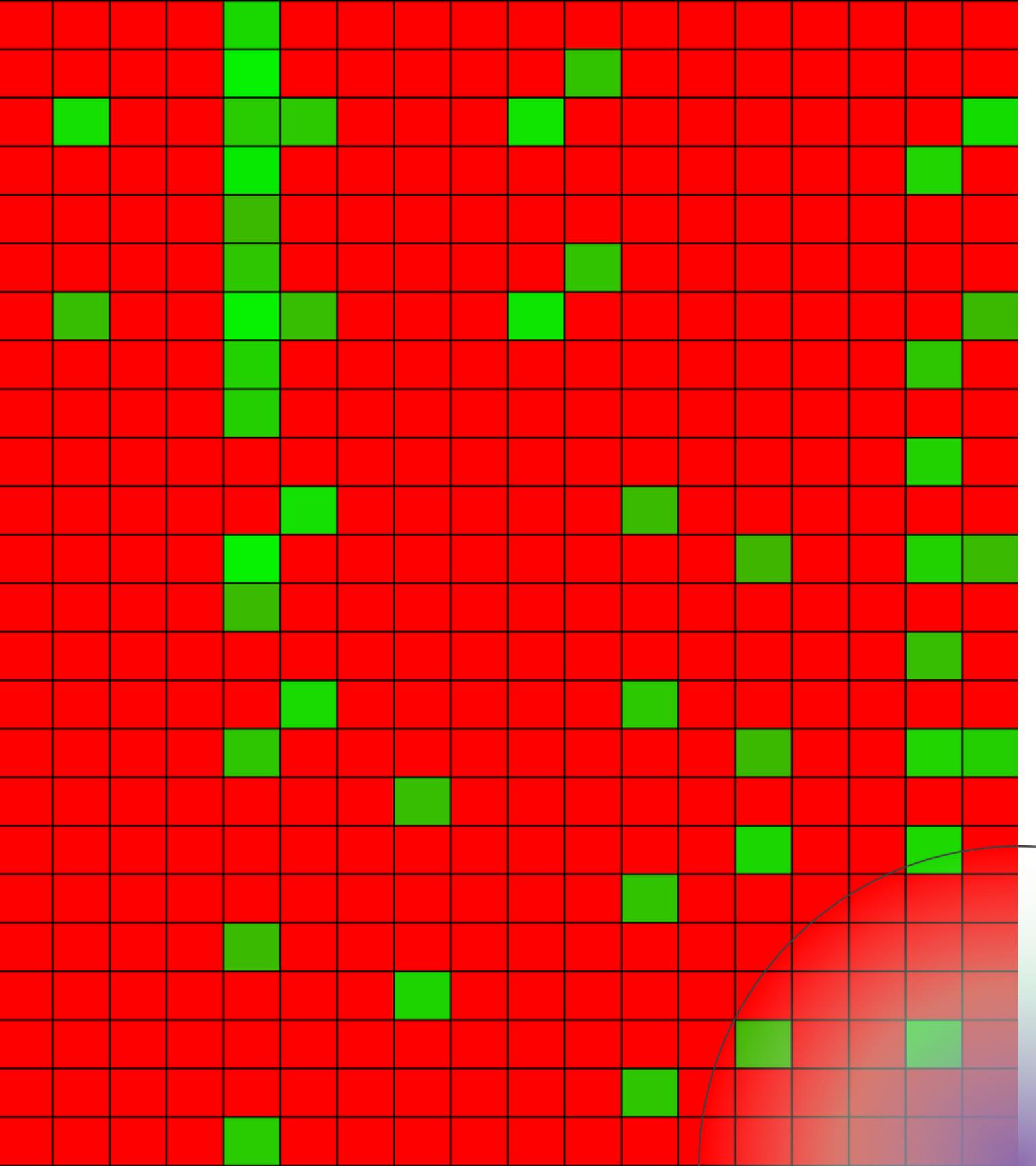
- build a **single convolutional neural network** for the whole period
- **compare it month by month** with the monthly-based ELMs

The reason why is to **check** if **extending the area of the local variables** also in the surroundings of Rijnland could bring a **more exhaustive local context** to the CNN which can turn into a **better bridging of Global and Local climate conditions** (crucial for sub-seasonal lead-times)

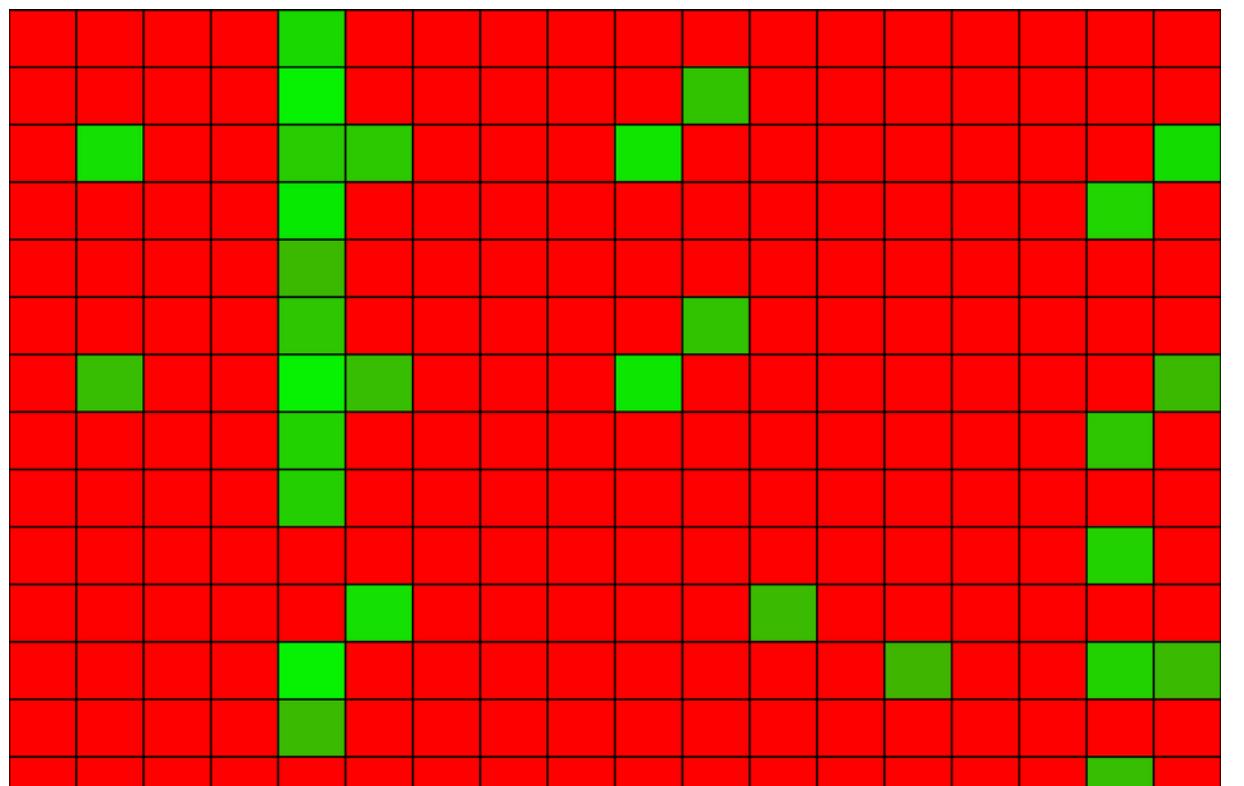
# Next steps

- 01 Neural Network
  - 02 Conv. Neural Net.
- 





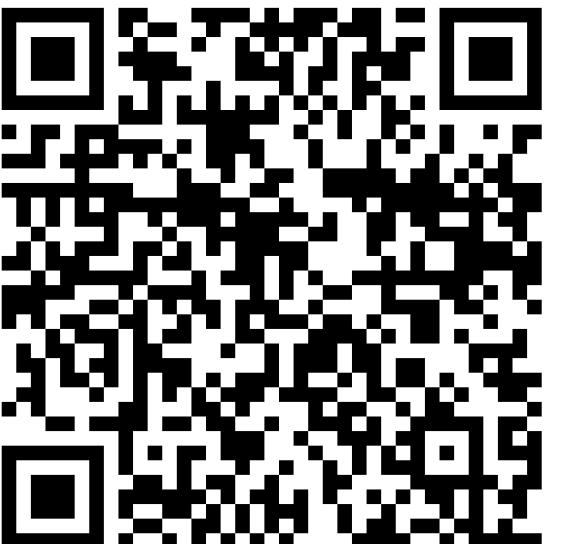
**Thank you  
for attending!**



Felsche et al. (2021)



Zimmerman et al. (2016)



Giuliani et al. (2019)



Our readaptation



you can find  
the slides  
here!