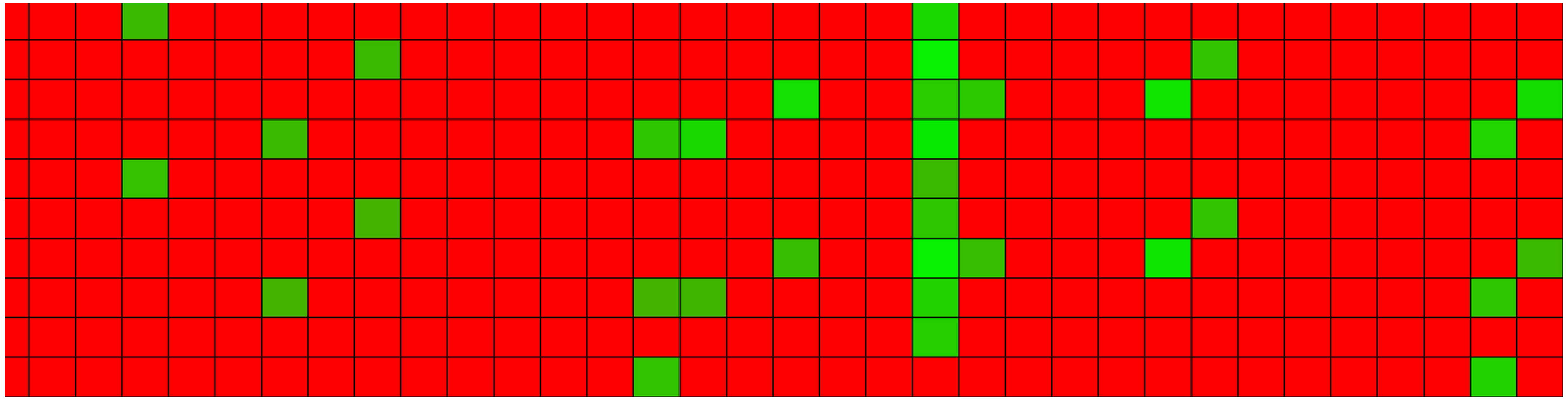
Improving sub-seasonal drought forecasting with machine learning and climate indices

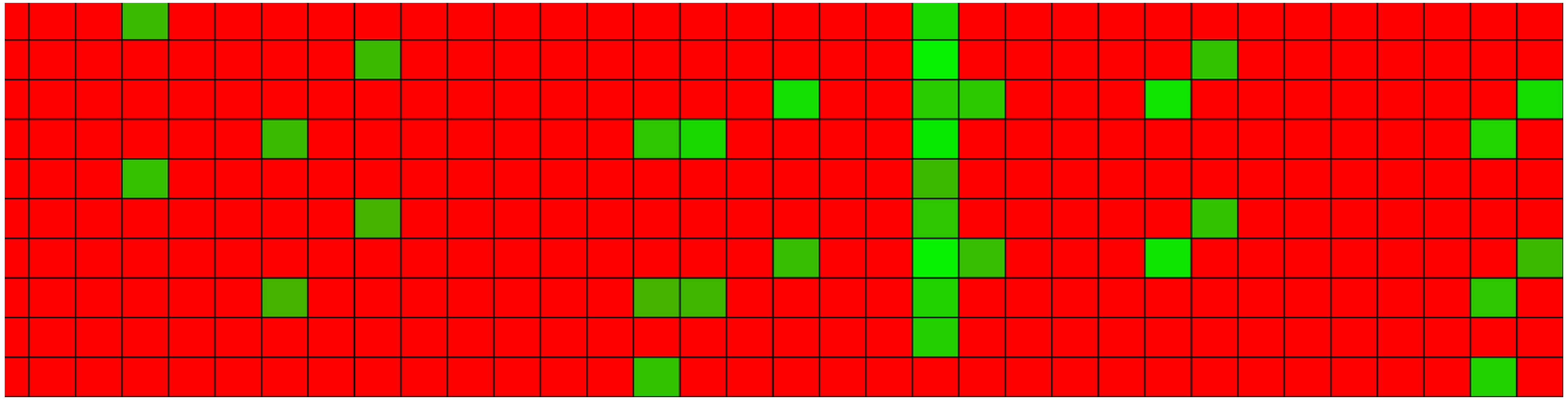
BOSSO FRANCESCO

IHE Delft, 08/12/2022



you can find the slides here!

BOSSO FRANCESCO



IHE Delft, 08/12/2022

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

Meteorological Drought

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

Reduced soil moinsture,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

Earth observation data
Artificial Intelligence
Hardware (GPU,TPU)

→ AI-based prediction models

McGovern et al. (2017)

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

Intro

Why to focus on sub-seasonal lead times?

Informative predictors

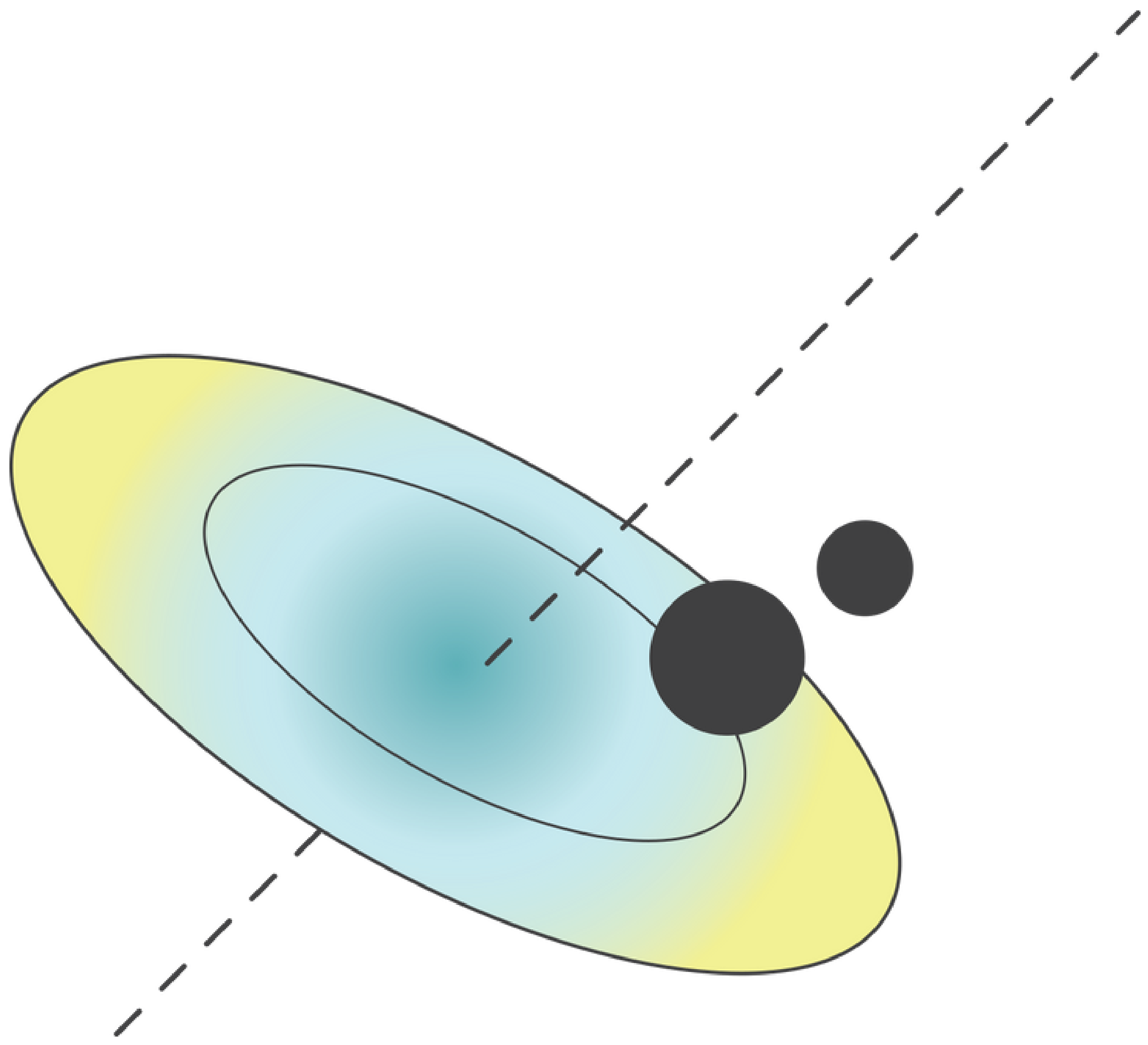
seasonal:
climate indices and large scale teleconnection patterns

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

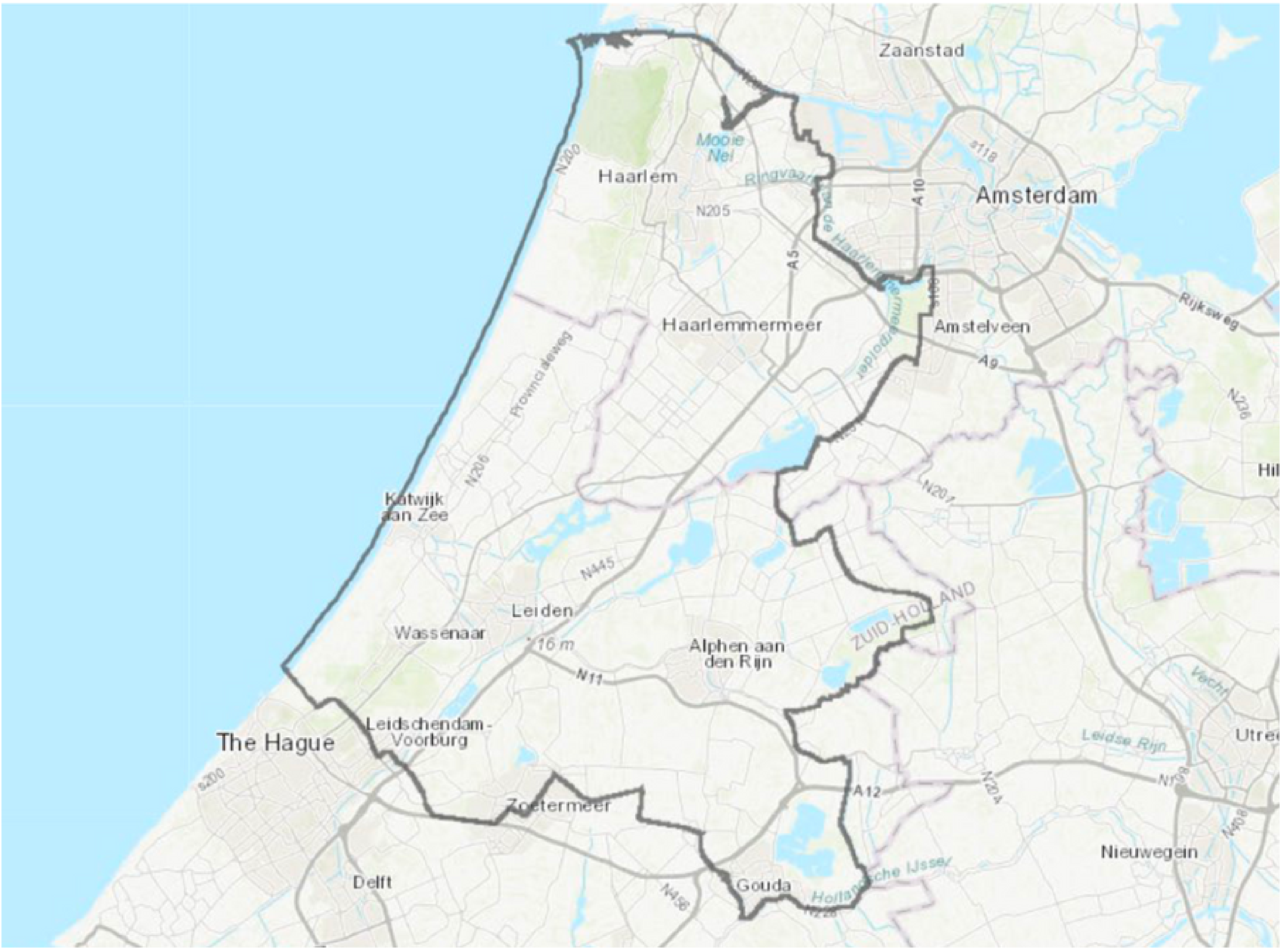
- **01 What (our goal)**
- 02 Where (study area)
- 03 How (the framework)

Machine Learning model for sub-seasonal precipitation forecasting

precipitation forecasting → SPI → drought forecating

Context

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



Rijnland

small sub-catchment of 1000 km2 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

Nino Index Phase Analysis (NIPA)

Zimmerman et al. (2016)

↓

Giuliani et al. (2019)

↓

Our readaptation

\+

Neural Networks

Neural Network

Convolutional Neural Network

Extreme Learning Machine

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

climate indices

El Niño Southern Oscillation (ENSO)

North Atlantic Oscillation (NAO)

SCAndinavian oscillation (SCA)

East Atlantic oscillation (EA)

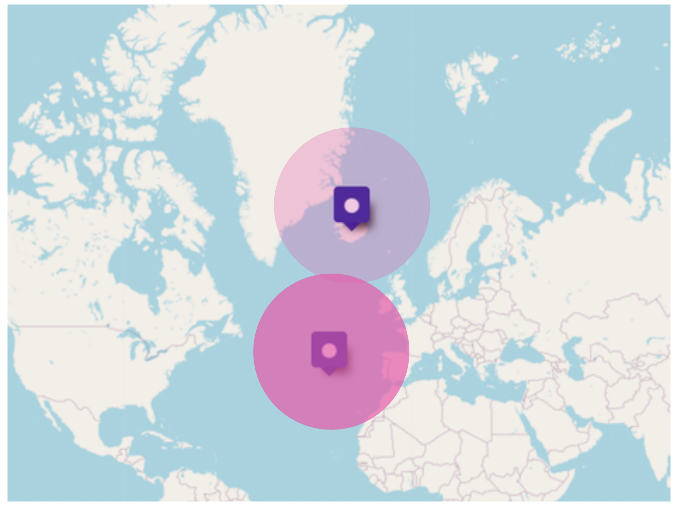
Framework

- 01 NIPA
- 02 ELM

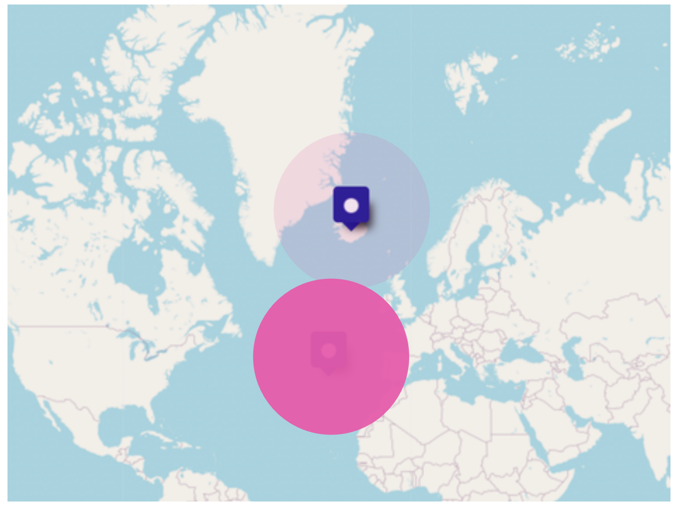
- above/below-normal temperatures in eastern United States and northern Europe
- above/below-normal temperatures in Greenland and southern Europe
- above/below-normal precipitation over northern Europe and Scandinavia
- above/below-normal precipitation over southern and central Europe

climate indices

North Atlantic Oscillation (NAO)



Phase Neg



Phase Pos

Framework

DATA

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

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
- Aggregation level 1

local precipitation of January and the global variable of December

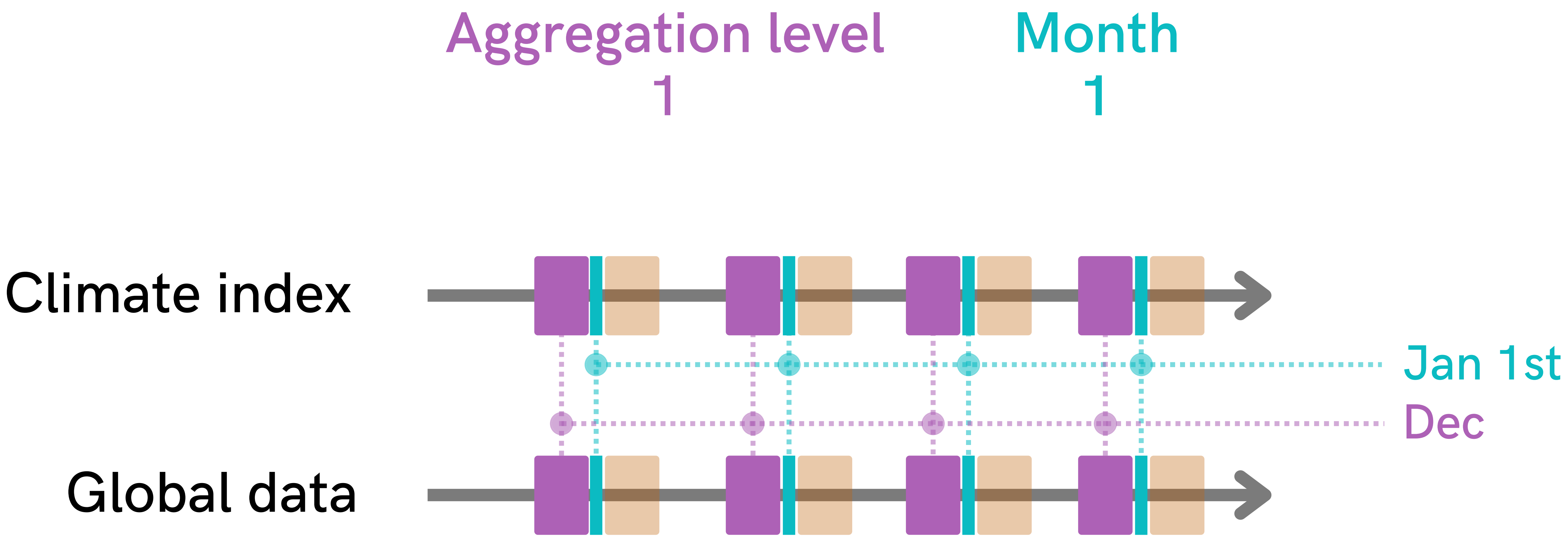
- Month 1
- Aggregation level 2

local precipitation of January and the global variable of November + December

Framework

● 01 NIPA
● 02 ELM

Input
Data extraction
Phase segmentation
Correlation
PCA
output

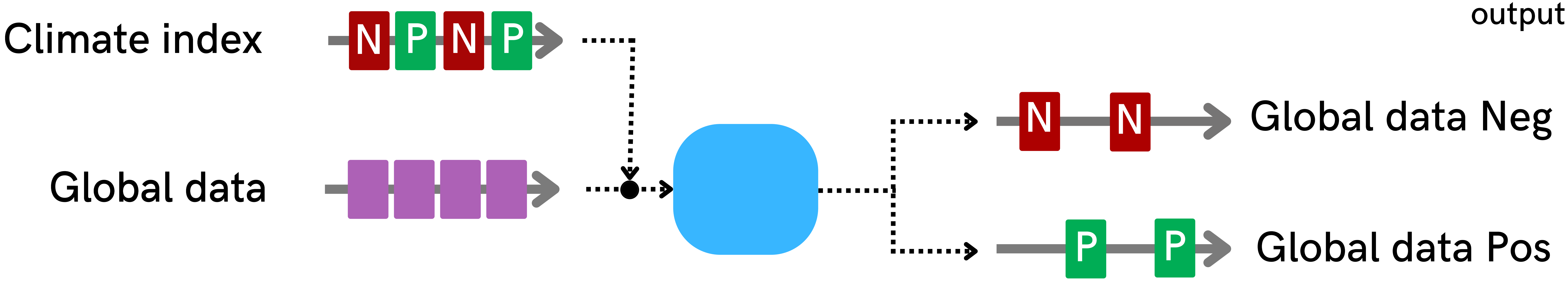


NOTE: this is an year-based operation. NIPA will extract the data for the December of each year

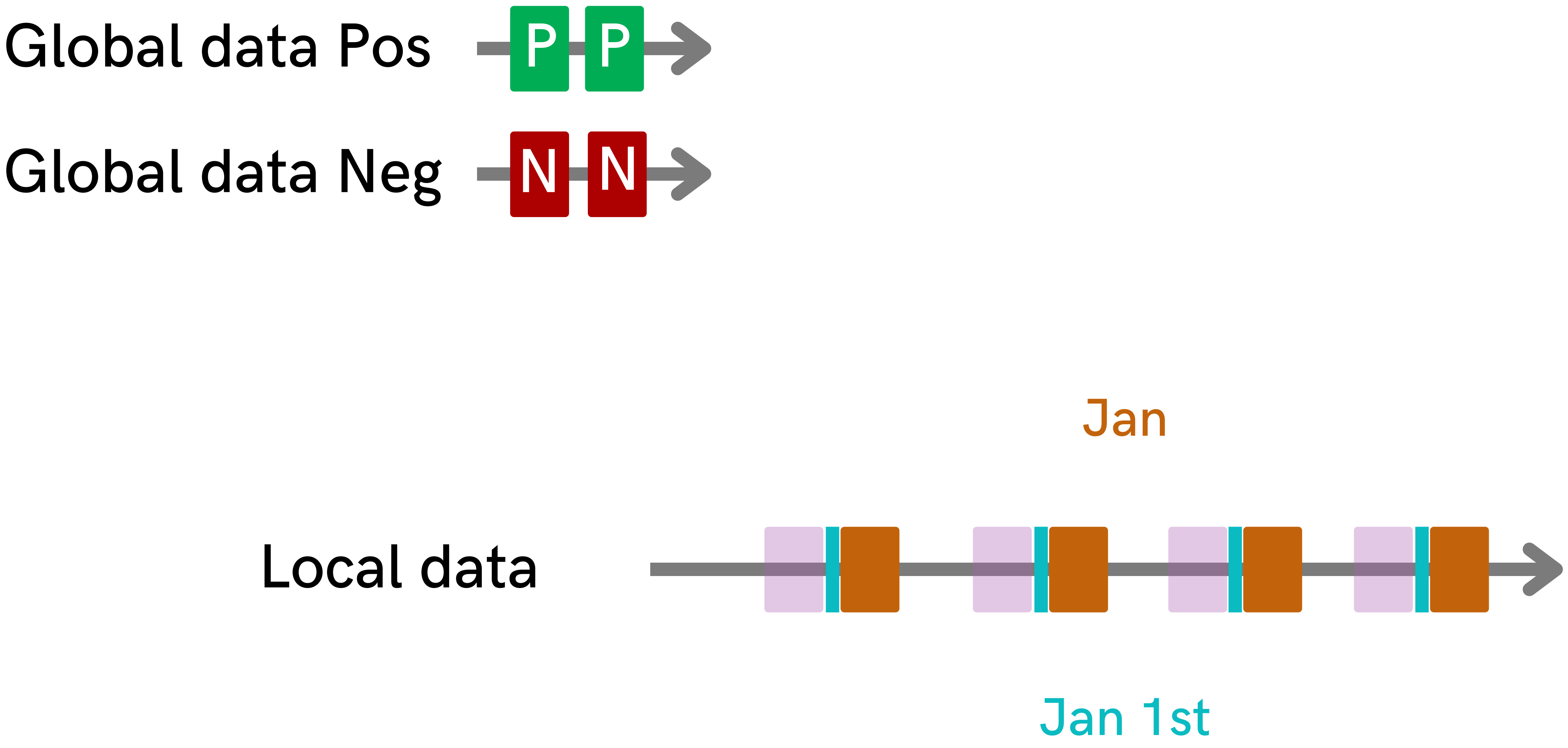
Framework

- 01 NIPA
- 02 ELM

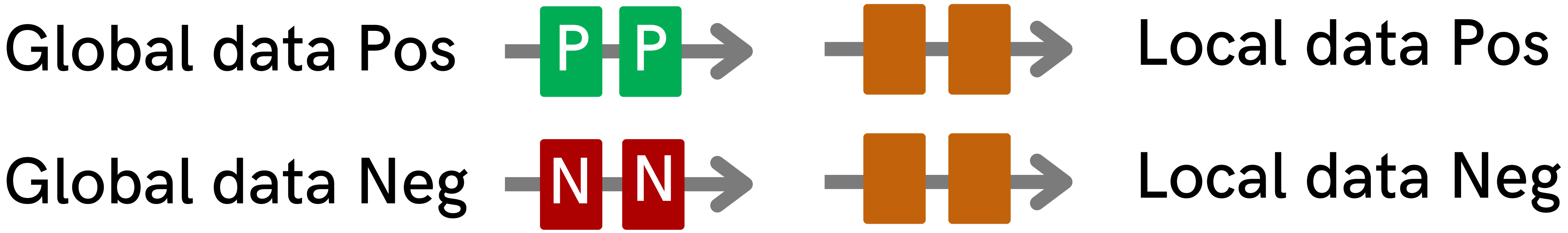
Input
Data extraction
Phase segmentation
Correlation
PCA
output



Framework



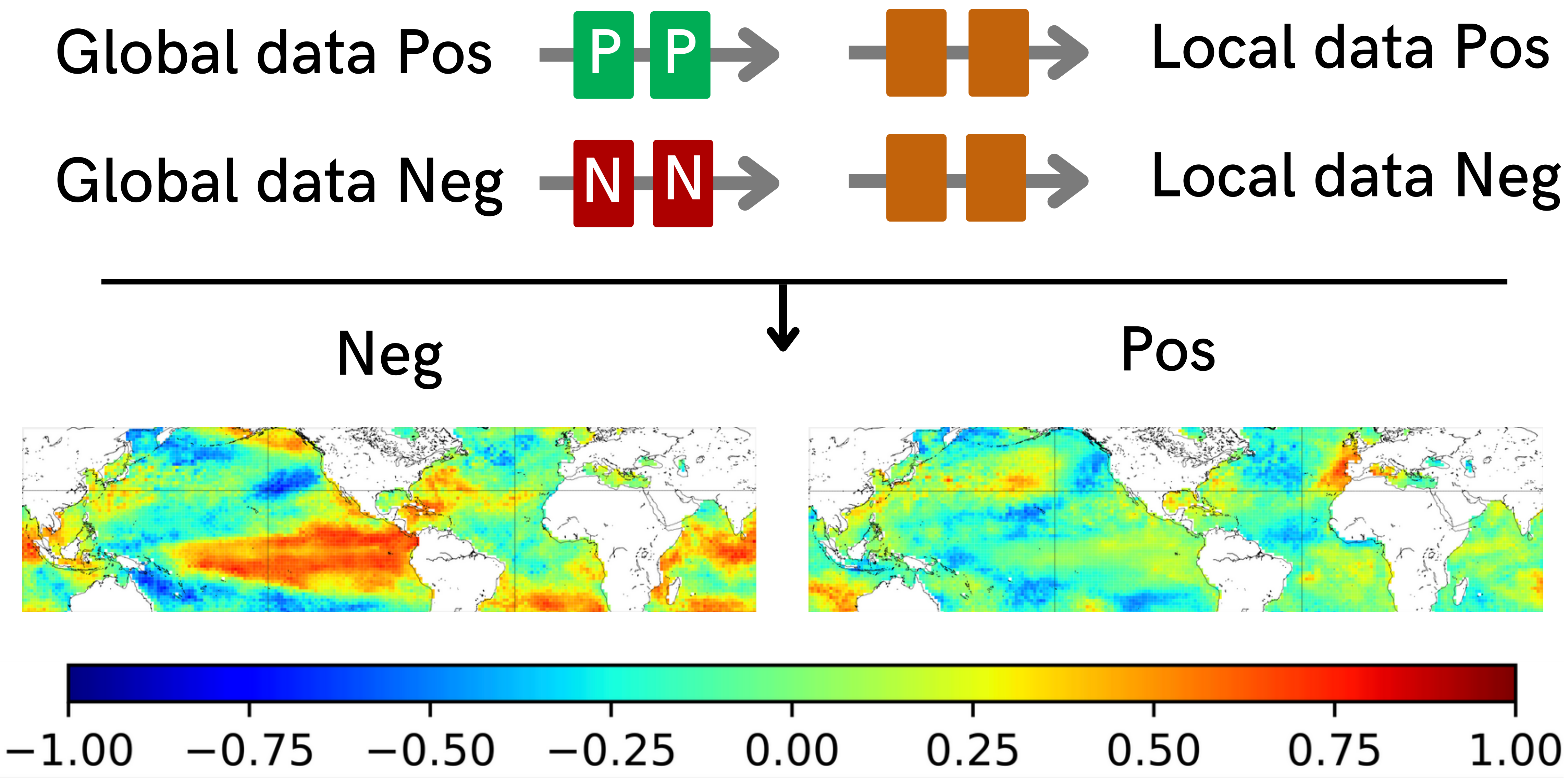
Framework



01 NIPA
02 ELM

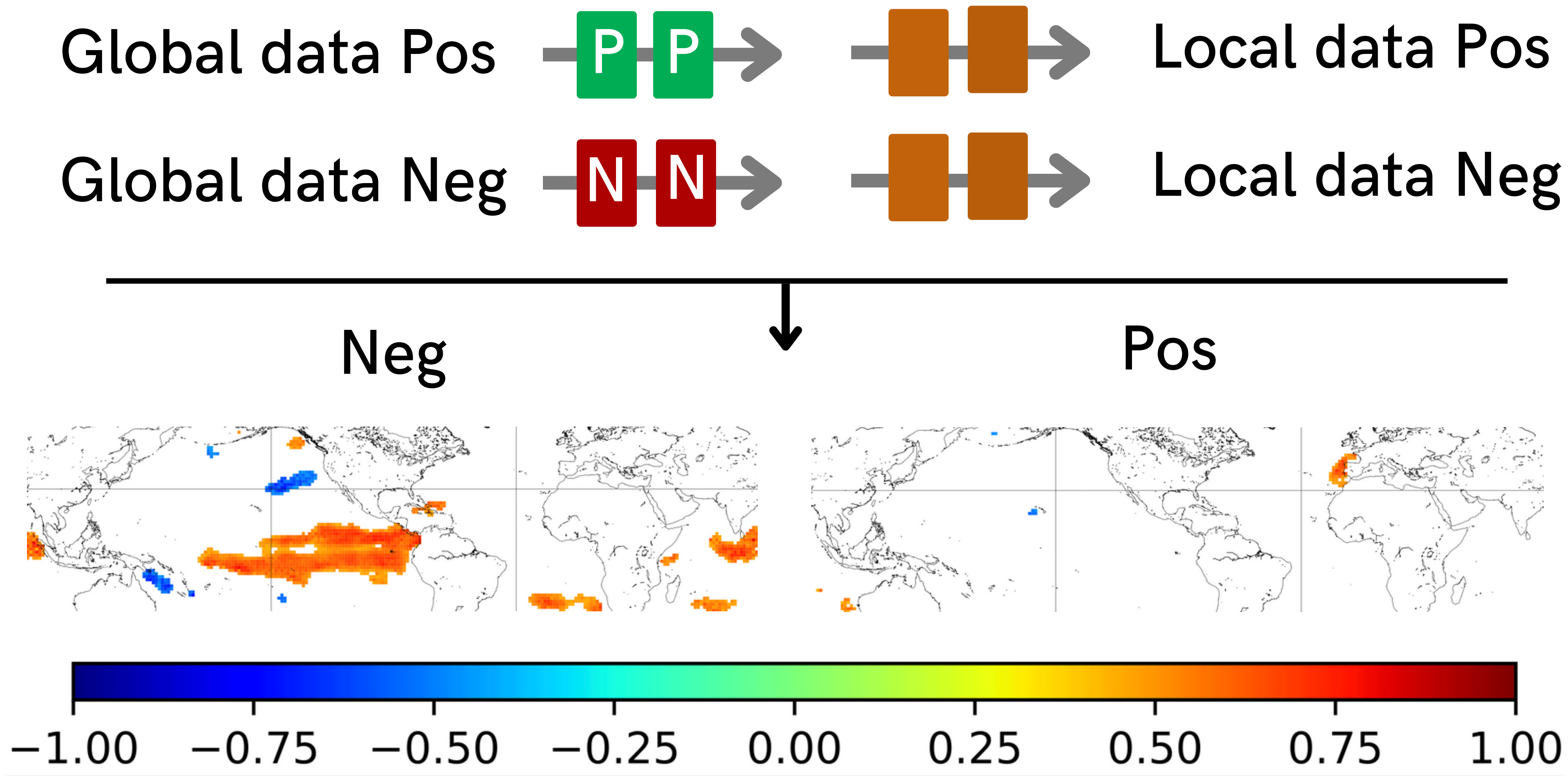
Input
Data extraction
Phase segmentation
Correlation
PCA
output

Framework



Correlation maps

Framework



95% of significance

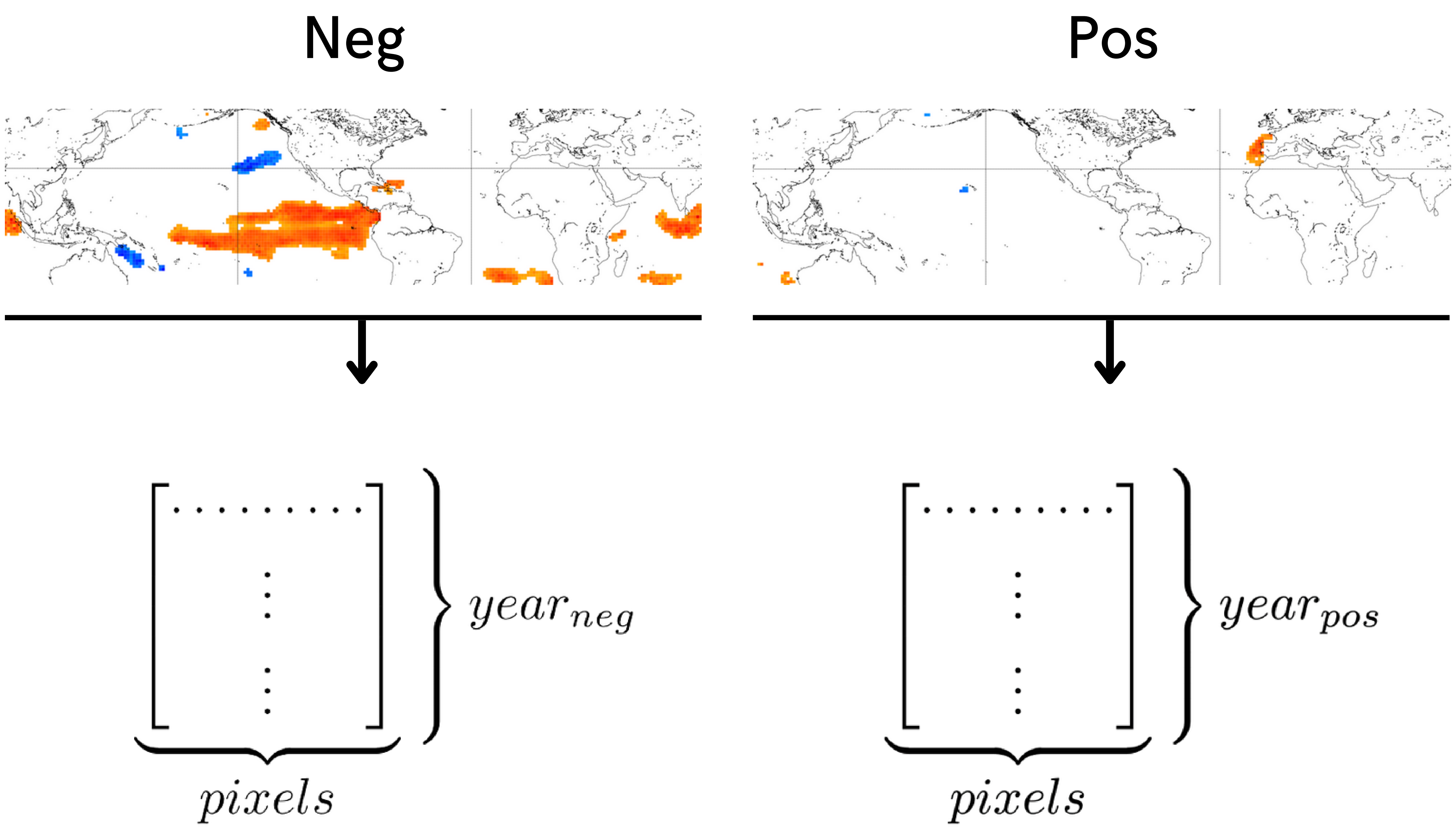
+

minimum correlation threshold 0.6

+

3x3 contiguous area check

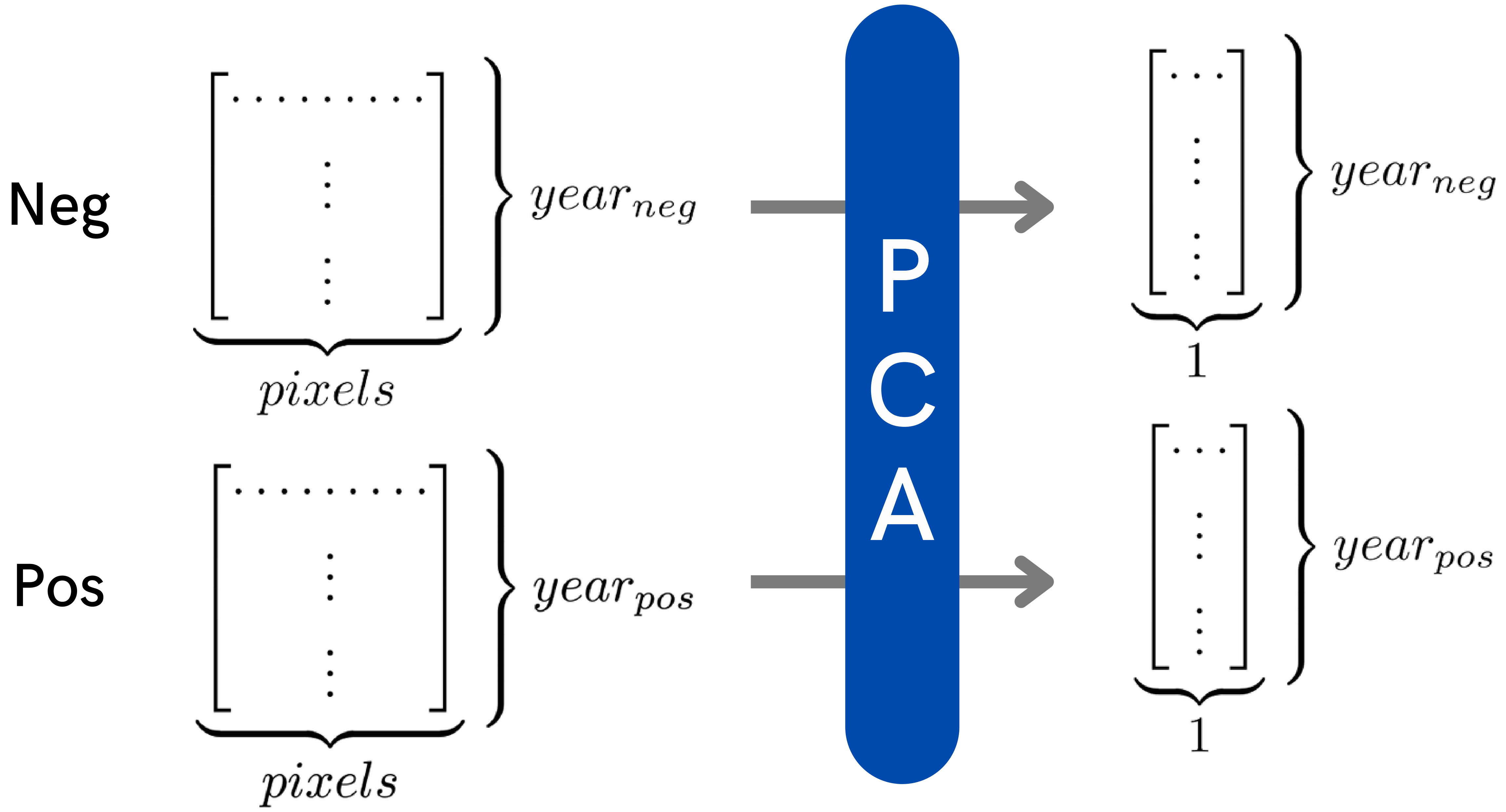
Framework



Framework

- 01 NIPA
- 02 ELM

Input
Data extraction
Phase segmentation
Correlation
PCA
output



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

Framework

- 01 NIPA
- 02 ELM

	Jan			Feb			Mar			Apr			May			Jun		
	SST	MSLP	Z500	SST	MSLP	Z500	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	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	-0,66	0	0	0	0	0	0	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_N	0	0	0	0	0	0	0	0	-0,67	0	0	0	0	0	0	0	-0,71	0,75
SCA_P	0	0	0	0	0	-0,69	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	-0,64	0	0	0	0	0	0	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_P	0	0	0	0	0	0	0	0	-0,64	0	0	0	0	0	0	0	-0,64	-0,65

	Jul			Aug			Sept			Oct			Nov			Dec		
	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500
SCA_N	0	0	0	0	0,75	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0,84	0	0	0	0	0	-0,7	0	0	0	0	0	0	0
ENSO-mei_N	0	0,78	0	0	0,71	0,7	0	0	0	-0,82	0	0	0	0	0	0	0	-0,79
NAO_N	0	0	0	0	0,81	0	0	0	0	0	0	0	0	0	0	0	0,74	0
SCA_P	0	0	0	0	-0,66	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0,69	0	0	0	0	0	0,68	0	0	0	0	0	0	0
ENSO-mei_P	0	0,66	0	0	-0,87	0,66	0	0	0	0,8	0	0	0	0	0	0	0	-0,66
NAO_P	0	0	0	0	0,73	0	0	0	0	0	0	0	0	0	0	0	0,69	0

	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500
SCA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,7	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,62	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500
SCA_N	0	0	0	0	0,72	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,73	0
ENSO-mei_N	0	0	0	0	0	0,78	0	0	0	0	0	0,65	0	0	0	0	0	0
NAO_N	0	0	0	0	0,84	0	0	0	0	0	0	0	0	-0,65	0	0	0,73	-0,67
SCA_P	0	0	0	0	0,65	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,66	0
ENSO-mei_P	0	0	0	0	0	0,76	0	0	0	0	0	0,7	0	0	0	0	0	0
NAO_P	0	0	0	0	-0,71	0	0	0	0	0	0	0	0	-0,66	0	0	0,74	0,72

	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500
SCA_N	0	0	0	0	0	0	0	0	0	0	-0,79	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	-0,7	0	0	0	0	0	-0,68	0,72
NAO_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCA_P	0	0	0	0	0	0	0	0	0	0	0,68	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,65	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	-0,76	0	0	0	0	0	-0,69	-0,63
NAO_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500	SST	MSLP	Z500
SCA_N	0	0	0	0	0	0	0	0,66	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,77	0	0	-0,78	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	-0,7	0	0	0	0	0	0
NAO_N	0	0	0	0	-0,67	0	0	0	0	0	0	0	0	0	0	0	0	0
SCA_P	0	0	0	0	0	0	0	-0,76	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0,63	0	0	0,74	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	-0,71	0	0	0	0	0	0
NAO_P	0	0	0	0	-0,73	0	0	0	0	0	0	0	0	0	0	0	0	0

Framework

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.

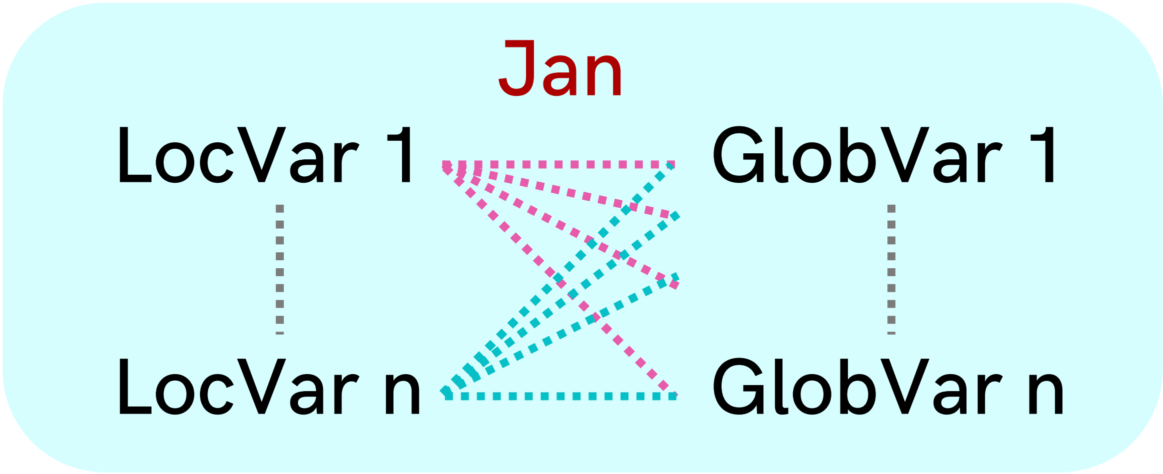
Framework

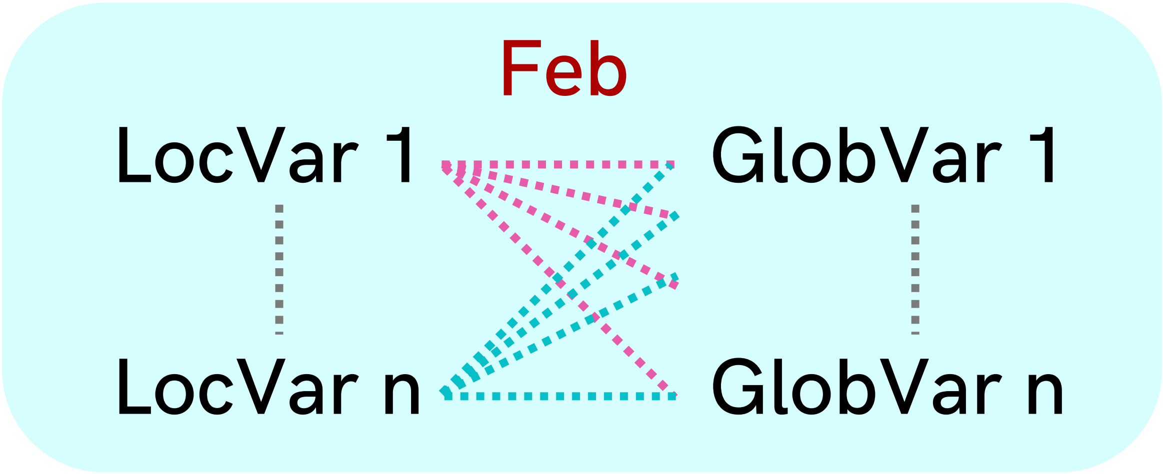
Creation of monthly-based datasets with all the possible combinations between local variables and global variables (if present).

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



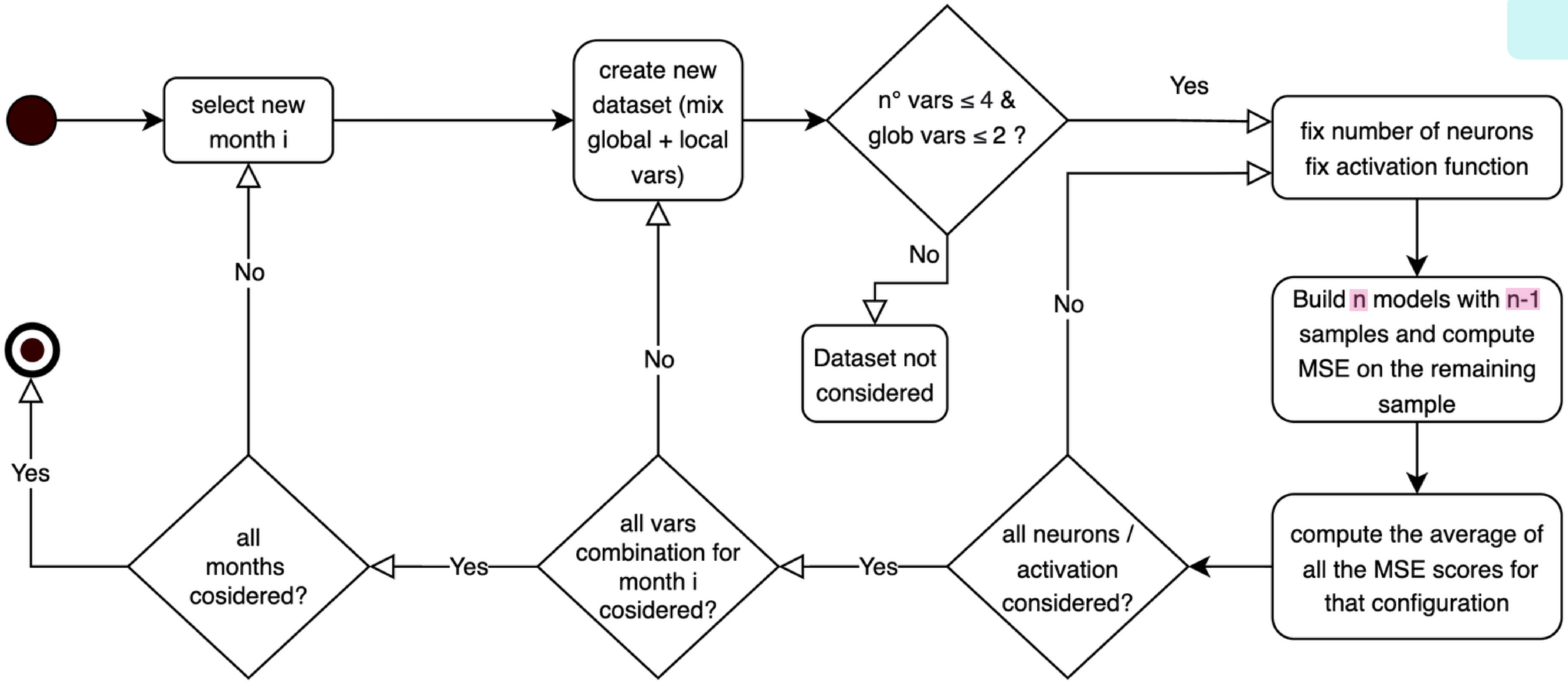


Framework

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*

Framework



n = total number of samples

Framework

- 01 NIPA
- 02 ELM

Local Data

Local/Global combinations

Model Selection

Final Results

Comparison

	1	2	3	4	5	6	7	8	9	10	11	12
t2m-TCC-TCWV-VW	939.6731465425995	2104.4558487862473	1140.6112601937127	1538.9290068931787	1149.259328894041	1328.6839479795044	2445.804287561171	1728.2951141978695	1900.737170821421	1426.3593654123172	880.5774160726029	829.50784563345
t2m-TCC-tp-UW	1021.1279812090271	890.5625842325519	1143.1628556419041	892.2994444345994	887.0173026175052	1240.0889732275584	2004.4711811586326	1686.749954411303	2484.6740753290146	1619.4711044073047	774.8820668389824	963.687598698996
t2m-TCC-tp-VW	1337.4750982313417	1010.8135083986966	986.6586062981585	935.8838511556661	1129.6628746238541	1075.8524606019369	1733.3778063102786	1814.2003606880482	3039.0928030104005	1592.6128142472649	757.4943494354072	997.6573795948194
t2m-TCC-UW-VW	1016.4377249875791	1179.4104893712233	996.6404753853362	916.3429663233621	772.8266438308959	1352.8489146986706	1778.1307618854366	1796.0389239571941	3417.290661210588	1365.8284413991958	765.7375747711916	889.4505330938105
t2m-TCWV-tp-UW	1019.8613996397796	1100.8499211170708	1196.1342833019276	1082.963060421151	958.2289762412412	1824.9358984269954	2347.048924640295	1811.086692333887	2128.46213095795	1396.70809512625	804.7163982519024	947.3689741258828
t2m-TCWV-tp-VW	1021.8460297658423	1110.9289313618526	1001.2600553490769	822.4301129980579	1044.693019236041	1544.6480208236487	2594.1977266652657	1728.4072220007583	1517.1208143997815	1448.1763265245008	729.0510088170072	867.4356196848047
t2m-TCWV-UW-VW	1274.1196807857561	1021.7307782967248	1312.7217217725354	932.7062621614916	891.8378961475914	1446.8366074564947	2289.4829851114578	1996.4662829257848	1864.4639683957528	1711.8055076470991	963.4209798080692	866.4187622397919
t2m-tp-UW-VW	1025.841762083675	1113.736050382477	1031.8348562391639	869.1092569913907	835.0164071889571	1215.5747267377587	1694.5510082885764	2128.3369495830957	2162.3475659166156	1783.8485606466234	761.4190743033507	1050.4542323474427
TCC-TCWV-tp-UW	951.0784956397018	1136.0035353389706	1097.6221479180986	951.5028547953407	978.9765194267907	1699.6895586695973	1572.1544580234847	1792.137437767528	1783.0723871313294	1356.9975240559454	862.2293451083755	1074.740870256806
TCC-TCWV-tp-VW	1111.631200059916	1045.3076485440868	1021.3863451148287	839.9186151975704	1173.8273458234178	1058.029824419771	1696.6376354432225	1802.70944231365	1681.6019364447898	1340.7818572528215	780.82105392201	749.8219069692552
TCC-TCWV-UW-VW	1100.2768107140803	998.1369621345418	933.7386791801532	922.7543183108378	1142.5035283520806	1352.4662405075358	1979.5141132935394	1897.9660903714048	1783.862839270157	1353.089055483501	777.6829842689885	858.3614540765325
TCC-tp-UW-VW	1077.7194249922006	1100.9102913951465	1172.066262479463	991.0788057439802	1096.8402758986263	1915.1914435822227	1576.0124346530172	1842.2406007732602	1899.7000936963045	1322.6637797279736	754.1409398513506	1212.3207281354012
TCWV-tp-UW-VW	931.371585333954	1057.4924991591058	998.5340448642792	891.5910342392315	1239.525262616883	1556.2849721357939	1720.5713677679028	2019.2842643557194	1567.582720183188	1548.540474357727	805.9897044922575	860.438944178655
SCA_Z500-1_tp-2_dataset.csv		642.2803411792628										
MER-SCA_Z500-1_tp-2_dataset.csv		608.7780644997616										
MSSHF-SCA_Z500-1_tp-2_dataset.csv		584.0578652692778										
RH-SCA_Z500-1_tp-2_dataset.csv		650.410992995989										
SD-SCA_Z500-1_tp-2_dataset.csv		605.6466203970995										
SH-SCA_Z500-1_tp-2_dataset.csv		835.0211159996643										
t2m-SCA_Z500-1_tp-2_dataset.csv		593.0563342451711										
TCC-SCA_Z500-1_tp-2_dataset.csv		642.3651917812329										
TCWV-SCA_Z500-1_tp-2_dataset.csv		686.9494887882515										
tp-SCA_Z500-1_tp-2_dataset.csv		686.8916760292254										
UW-SCA_Z500-1_tp-2_dataset.csv		592.86978455185										
VW-SCA_Z500-1_tp-2_dataset.csv		670.4377691710046										
MER-MSSHF-SCA_Z500-1_tp-2_dataset.csv		674.2480288369807										
MER-RH-SCA_Z500-1_tp-2_dataset.csv		656.0754984401302										
MER-SD-SCA_Z500-1_tp-2_dataset.csv		592.3140764545021										
MER-SH-SCA_Z500-1_tp-2_dataset.csv		626.760117053719										
MER-t2m-SCA_Z500-1_tp-2_dataset.csv		753.7117071405977										
MER-TCC-SCA_Z500-1_tp-2_dataset.csv		644.8367293792238										
MER-TCWV-SCA_Z500-1_tp-2_dataset.csv		801.6438690315448										
MER-tp-SCA_Z500-1_tp-2_dataset.csv		493.1392795814359										
MER-UW-SCA_Z500-1_tp-2_dataset.csv		664.9644805793855										
MER-VW-SCA_Z500-1_tp-2_dataset.csv		902.3629736494933										
MSSHF-RH-SCA_Z500-1_tp-2_dataset.csv		746.2517321008878										
MSSHF-SD-SCA_Z500-1_tp-2_dataset.csv		699.8873520893462										
MSSHF-SH-SCA_Z500-1_tp-2_dataset.csv		515.2433239383465										
MSSHF-t2m-SCA_Z500-1_tp-2_dataset.csv		694.0808768325446										

one table for each combination of neurons/activation functions

22 tables

Framework

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-mei_MSLP-1_tp-7_dataset.csv'),
8: (380.20, 9, 'relu', 'SH-tp-SCA_MSLP-2_tp-8_dataset.csv-ENSO-mei_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-mei_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')

Framework

01 NIPA

02 ELM

Local Data

Local/Global combinations

Model Selection

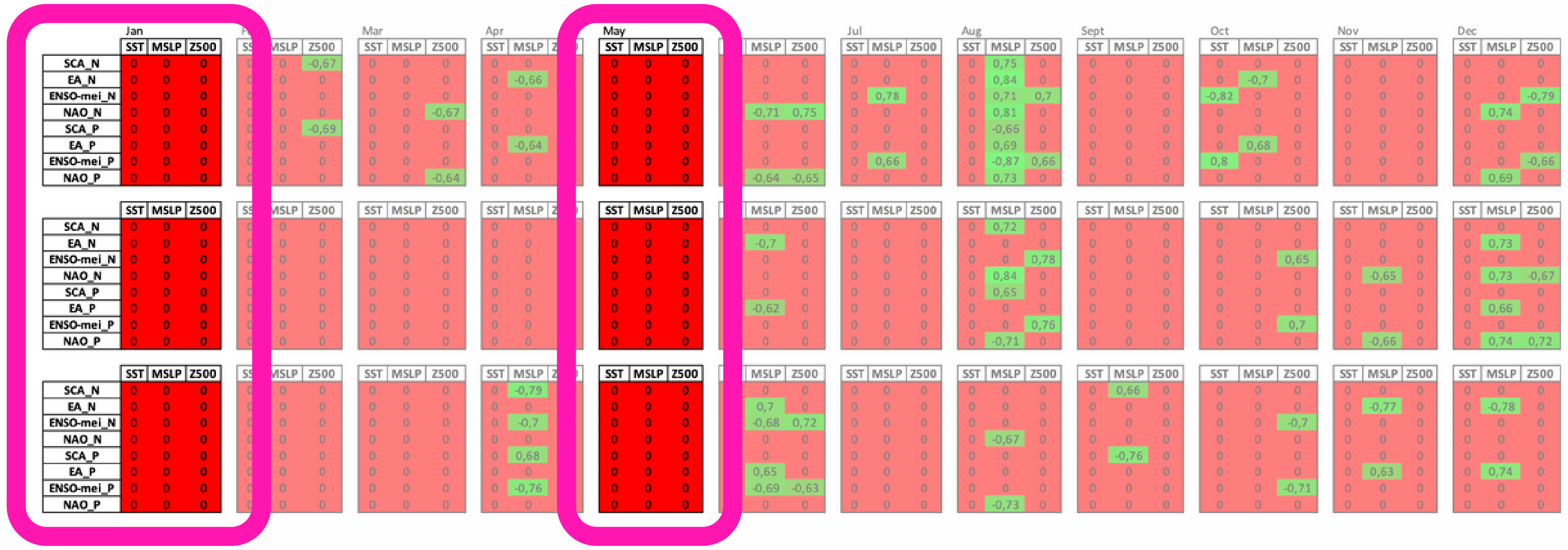
Final Results

Comparison

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

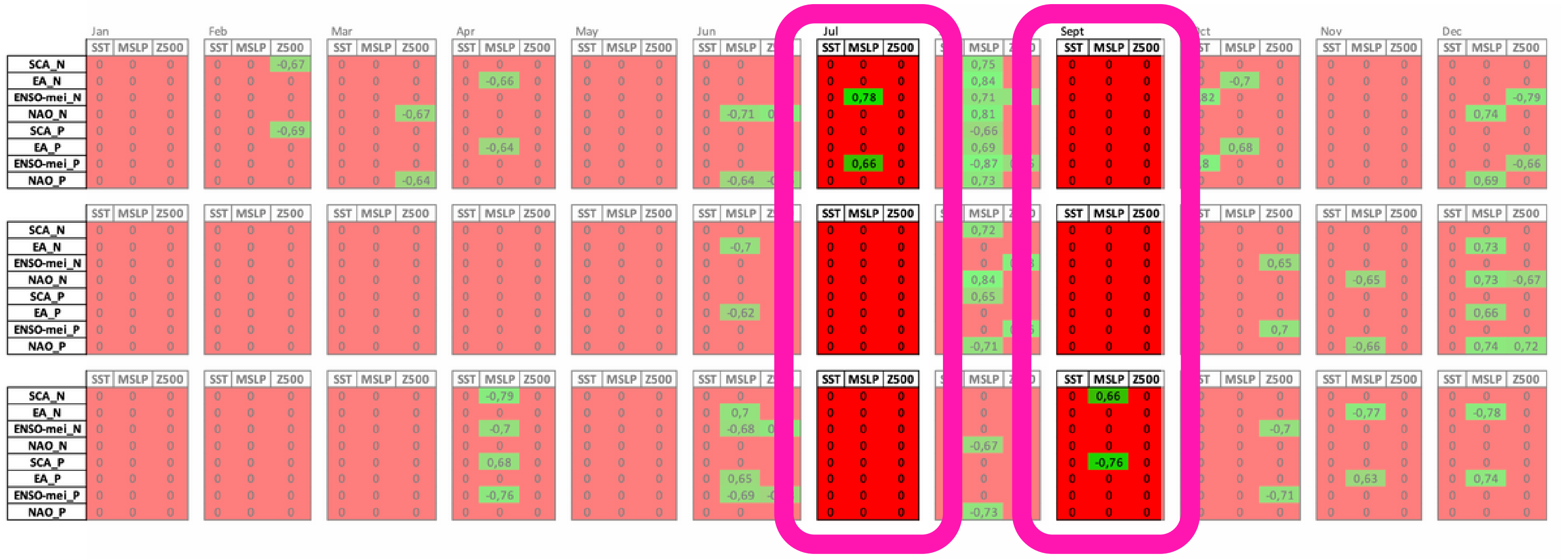


Framework

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



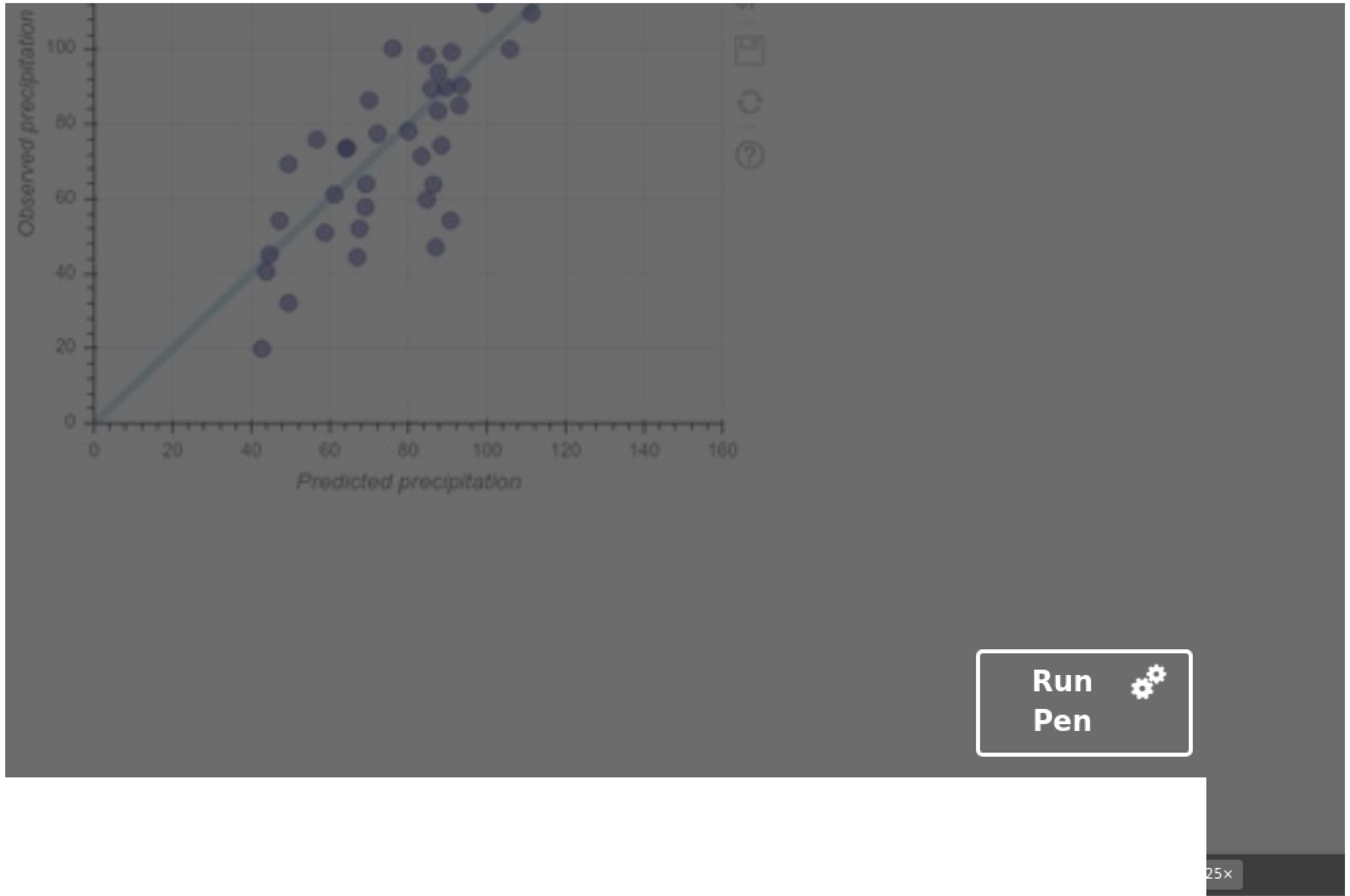
Framework

01 NIPA

02 ELM

- Local Data
- Local/Global combinations
- Model Selection
- Final Results
- Comparison

Build a model for each moth based on the best setting



Framework

Giuliani et. al.

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

Pearson prediction vs target: 0.81

Our work

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

Pearson prediction vs target: 0.81/0.82

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)

Next steps

- 01 Neural Network
- 02 Conv. Neural Net.

Next steps

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

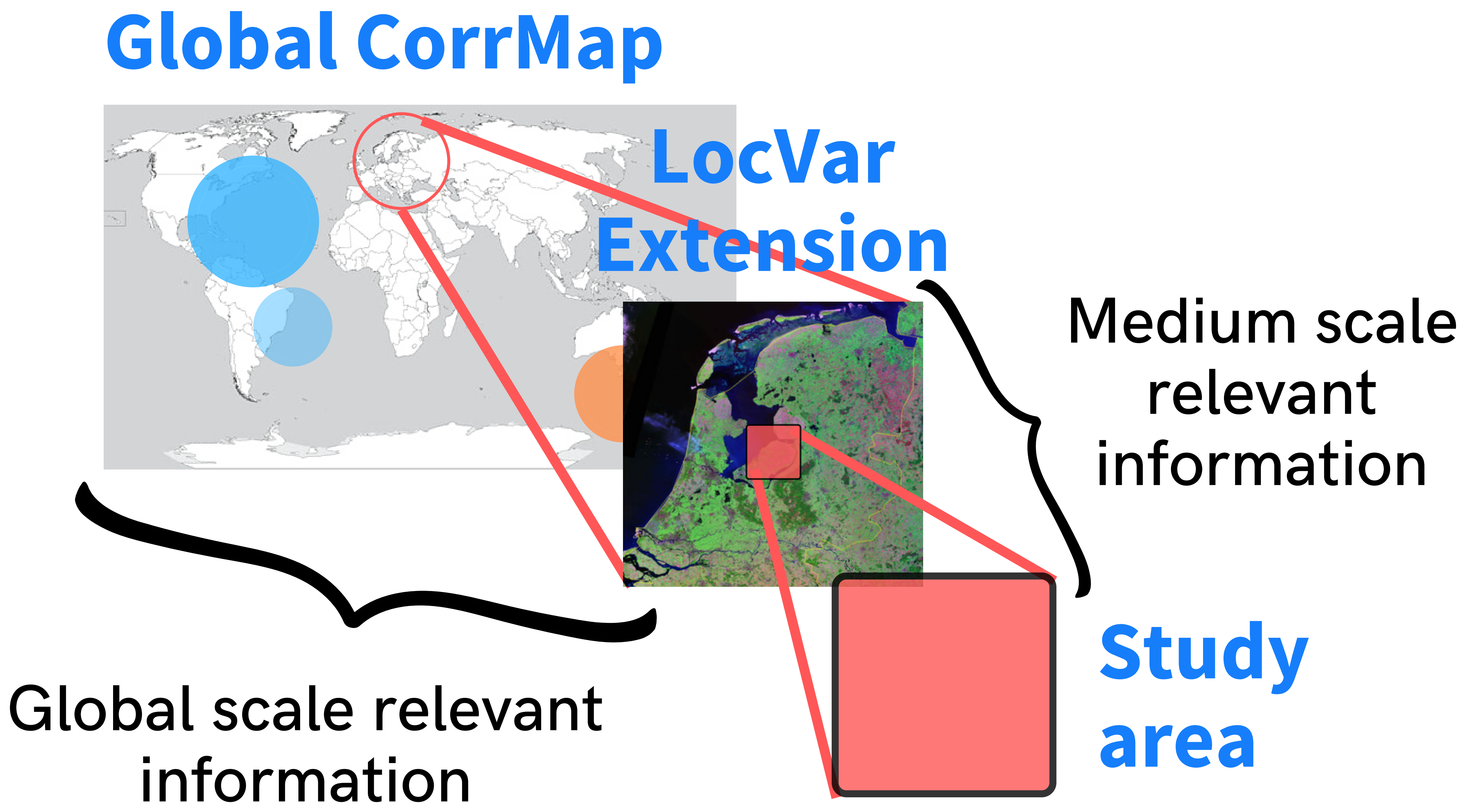
Next steps

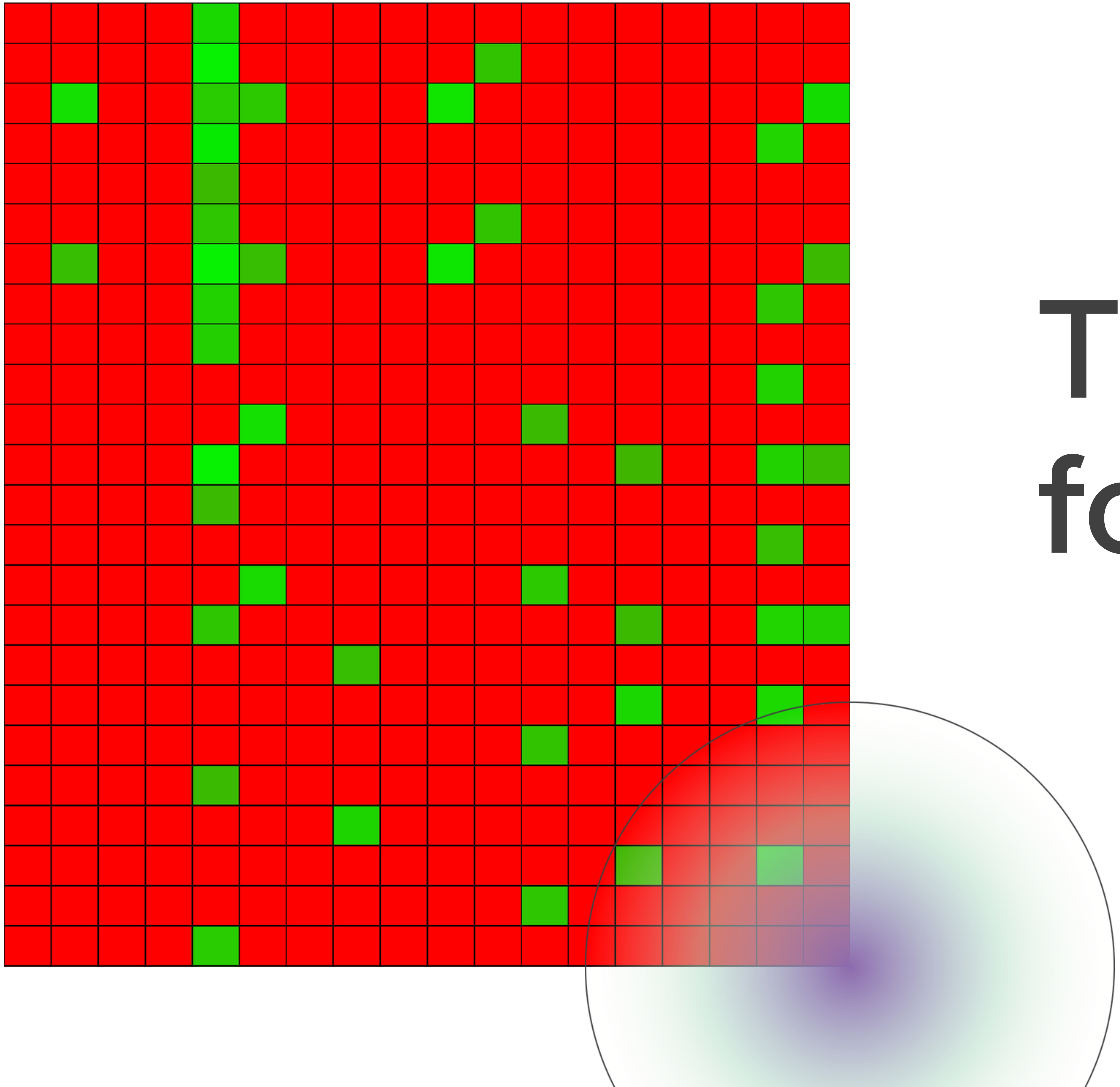
Because of the lack of samples and the absence of a proper testing procedure, we plan to:

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

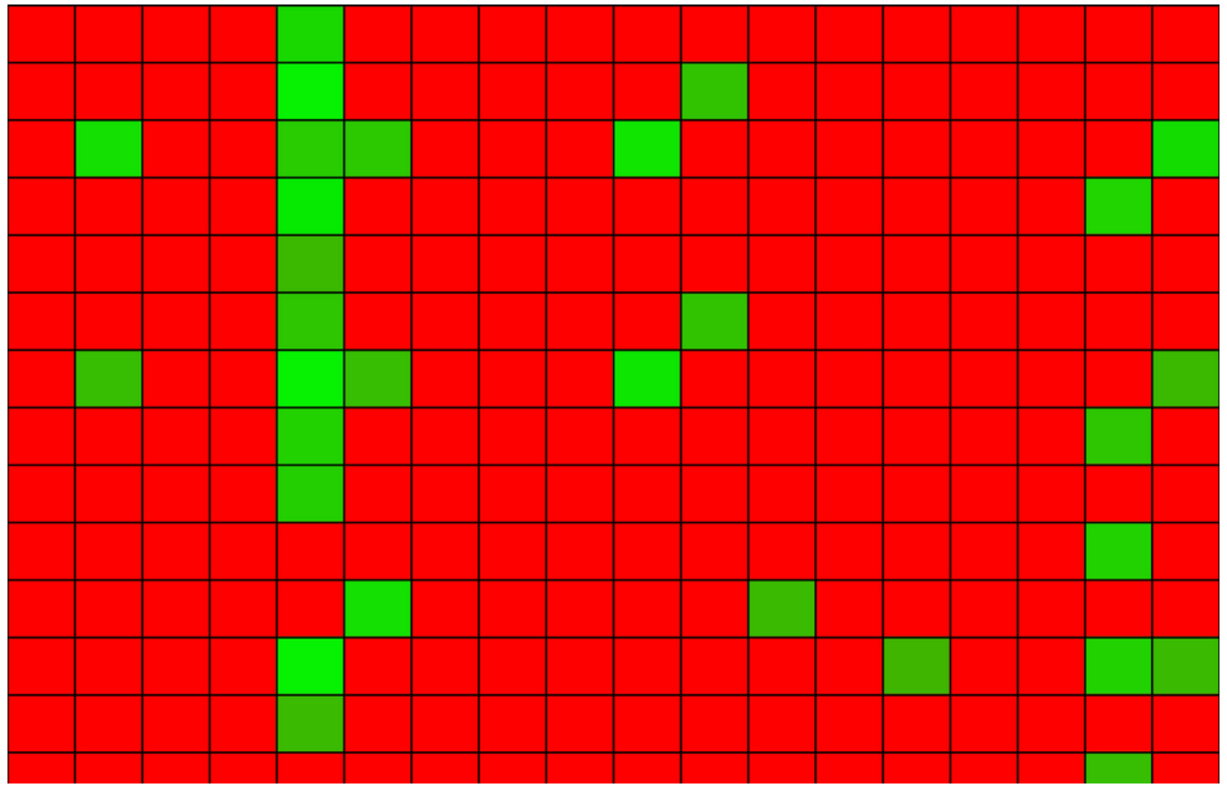
The reason why is to ***check*** if extending the area of the local variables also in the surroundings of Rijnland could bring a more exhaustive local context to the CNN which can turn into a better bridging of Global and Local climate conditions (crucial for sub-seasonal lead-times)

Next steps


Global CorrMap
LocVar Extension
Medium scale relevant information
Study area
Global scale relevant information



Thank you for attending!



you can find the slides here!

Felsche et al. (2021)

Zimmerman et al. (2016)

Giuliani et al. (2019)

Our readaptation