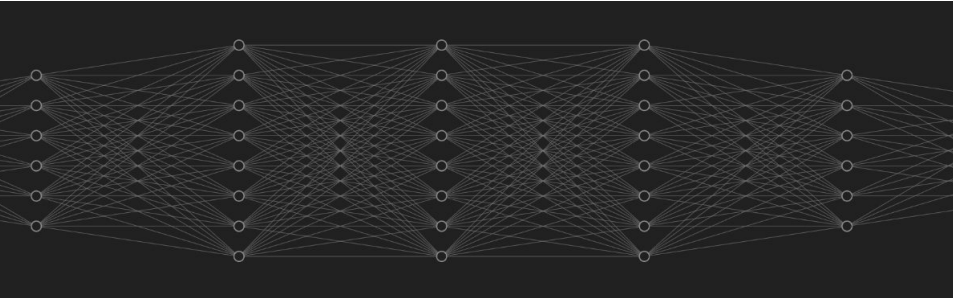


An Earth-system-oriented view of the S2S

predictability of North American weather regimes using ML



Jhayron Pérez-Carrasquilla and Dr. Maria Molina

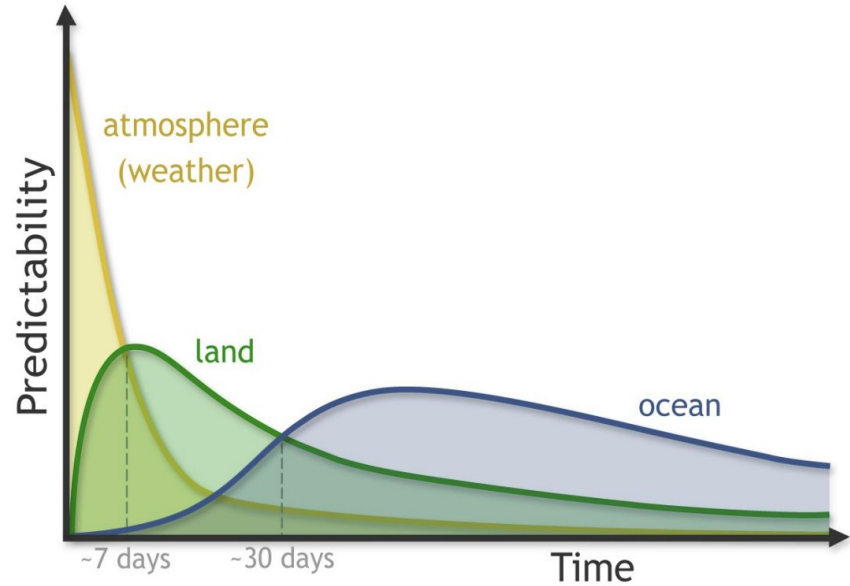
Department of Atmospheric and Oceanic Science

Research question

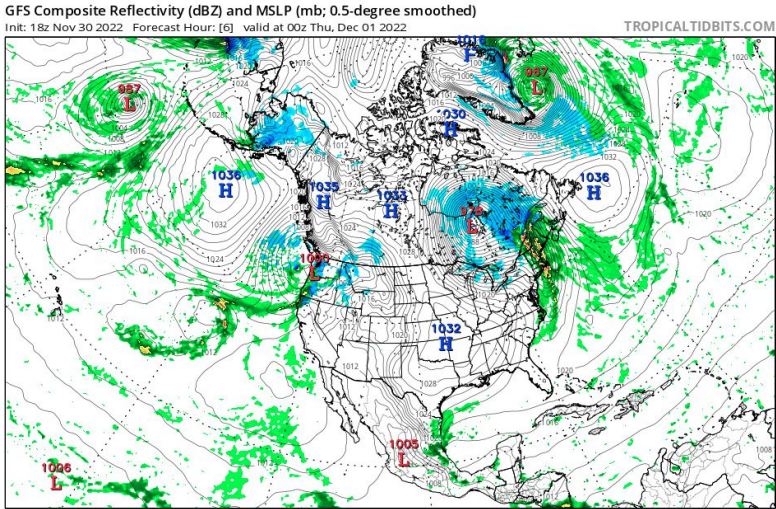
Where does atmospheric predictability come from?

What variables or Earth system components?

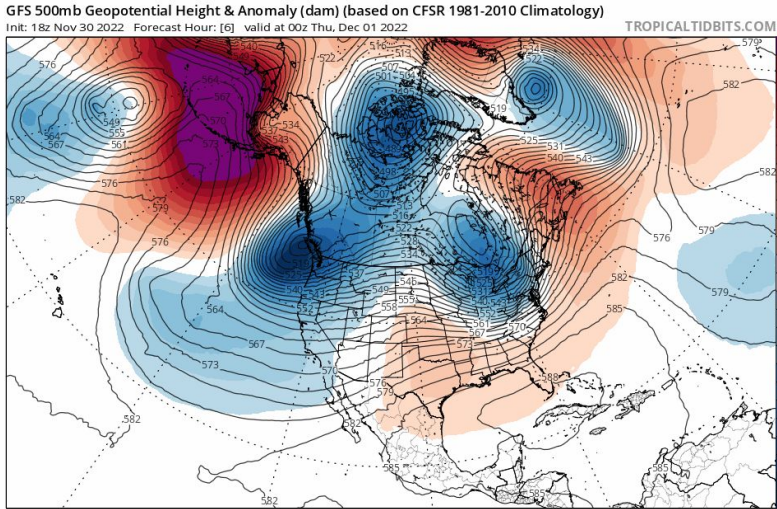
What processes?



What to predict? (How to "simplify" the atmosphere?)



Small-scale features



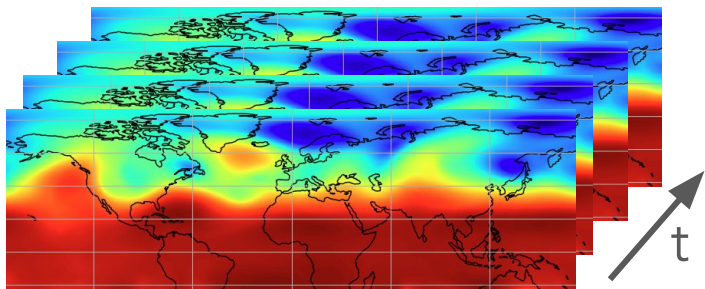
Large-scale features

Weather Regimes are easier to follow and understand and also affect the surface.

Images from: <https://www.tropicaltidbits.com/>

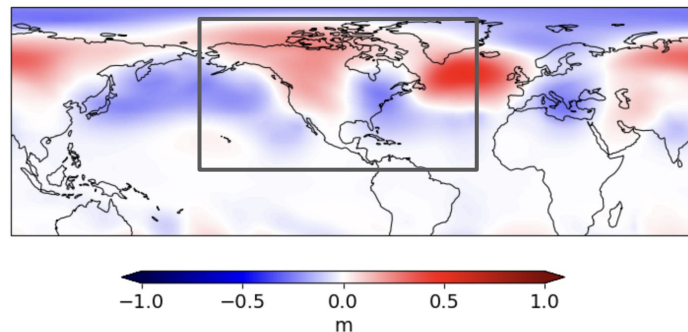
Computing year-round Weather Regimes

1) Bi-weekly 500hPa
Geopotential height (Z500)
1981-2020



2) Extracted region of interest, remove
annual cycle and linear trends

Z500 Anomalies

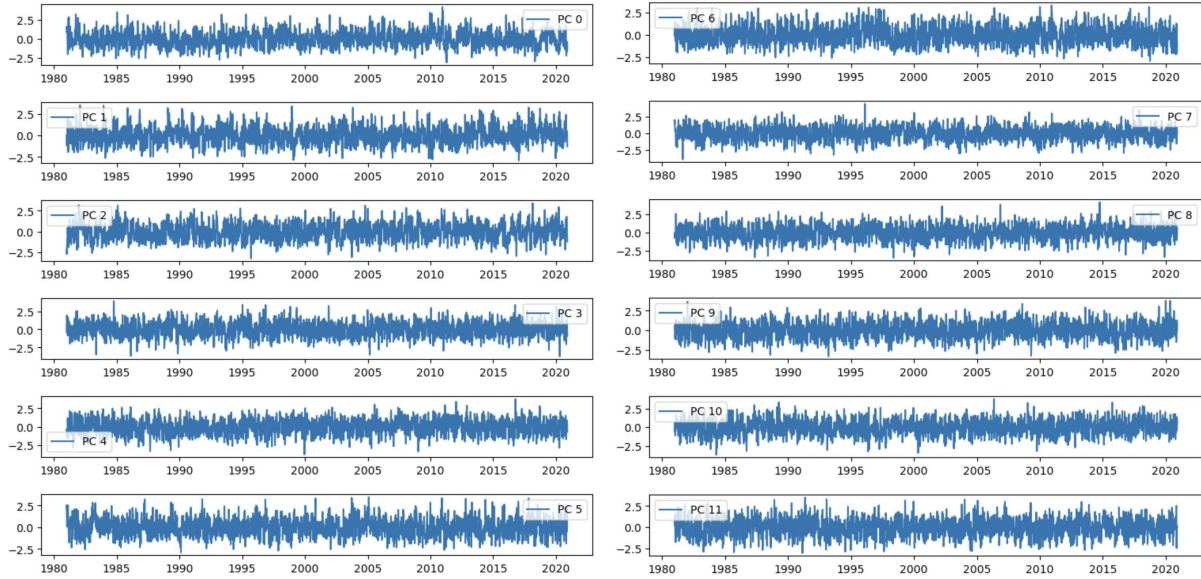


Molina, M. J., J. H. Richter, A. A. Glanville, K. Dagon, J. Berner, A. Hu, and G. A. Meehl, 2023. Subseasonal representation and predictability of North American weather regimes using cluster analysis. *Artificial Intelligence for the Earth Systems*, 1–54.

Lee, Simon H., Michael K. Tippett, and Lorenzo M. Polvani. "A new year-round weather regime classification for North America." *Journal of Climate* 36.20 (2023): 7091-7108.

Computing year-round Weather Regimes

3) Dimensionality reduction: 12 first PCs (85% of variance)



k-means clustering:
extract 4 categories

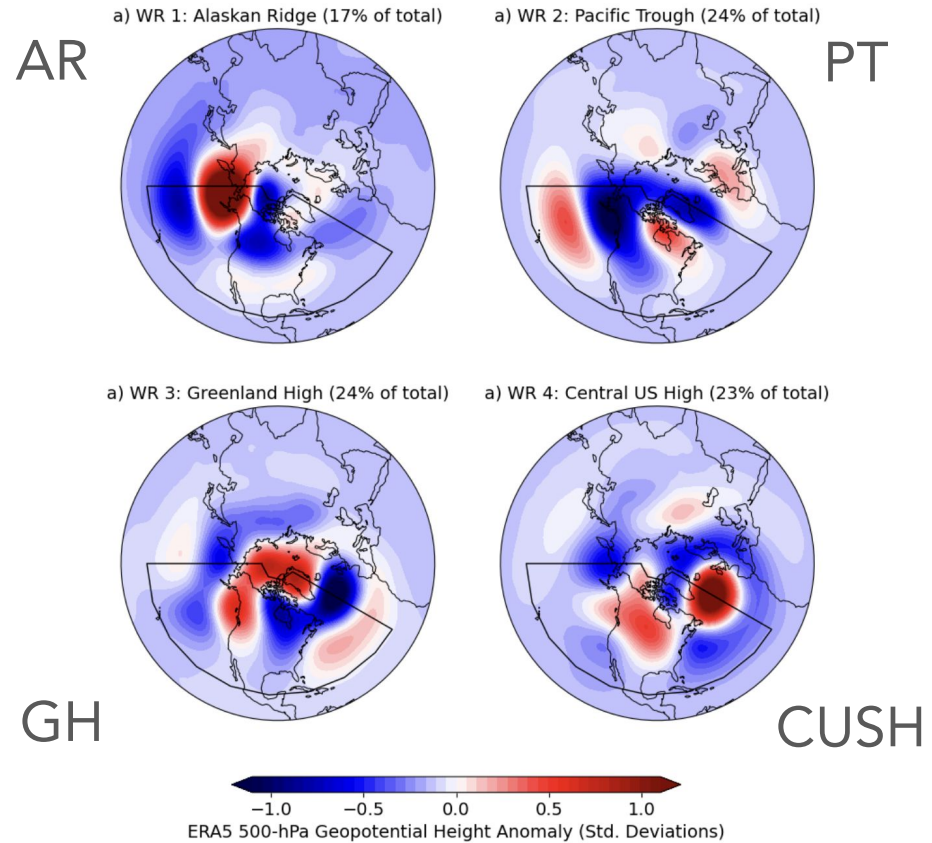


Molina, M. J., J. H. Richter, A. A. Glanville, K. Dagon, J. Berner, A. Hu, and G. A. Meehl, 2023. Subseasonal representation and predictability of north american weather regimes using cluster analysis. Artificial Intelligence for the Earth Systems, 1–54.

Lee, Simon H., Michael K. Tippett, and Lorenzo M. Polvani. "A new year-round weather regime classification for North America." Journal of Climate 36.20 (2023): 7091-7108.

Computing year-round Weather Regimes

4) 5 classes: 4 Weather regimes
+ 1 No WR class (near
climatology)



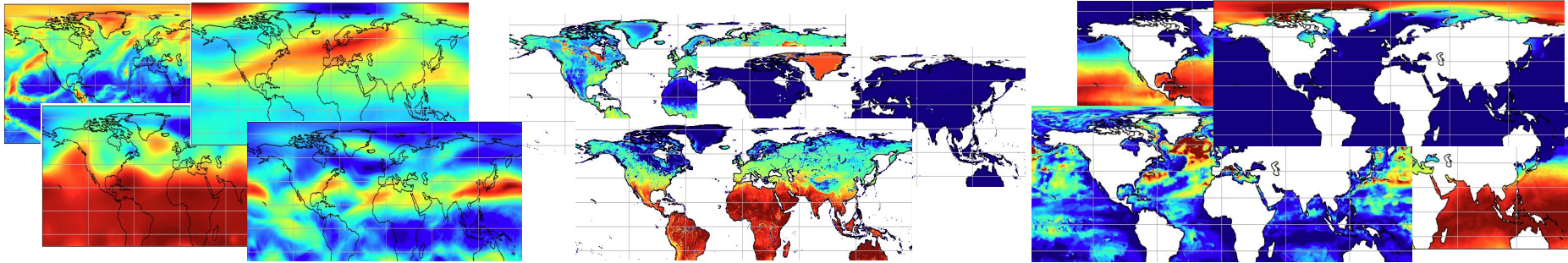
Molina, M. J., J. H. Richter, A. A. Glanville, K. Dagon, J. Berner, A. Hu, and G. A. Meehl, 2023. Subseasonal representation and predictability of north american weather regimes using cluster analysis. Artificial Intelligence for the Earth Systems, 1–54.
Lee, Simon H., Michael K. Tippett, and Lorenzo M. Polvani. "A new year-round weather regime classification for North America." Journal of Climate 36.20 (2023): 7091-7108.

Predictors representing the initial state of the Earth system

Atmosphere (ERA5) - 4 variables: Z500hPa, U10hPa and U200hPa, and OLR.

Land (ERA5) - 9 variables: Soil integrated moisture and heat for different depths, and snow depth.

Ocean (SODA) - 10 variables: OHC for different depths, SSH, SST, MLD, ice properties.



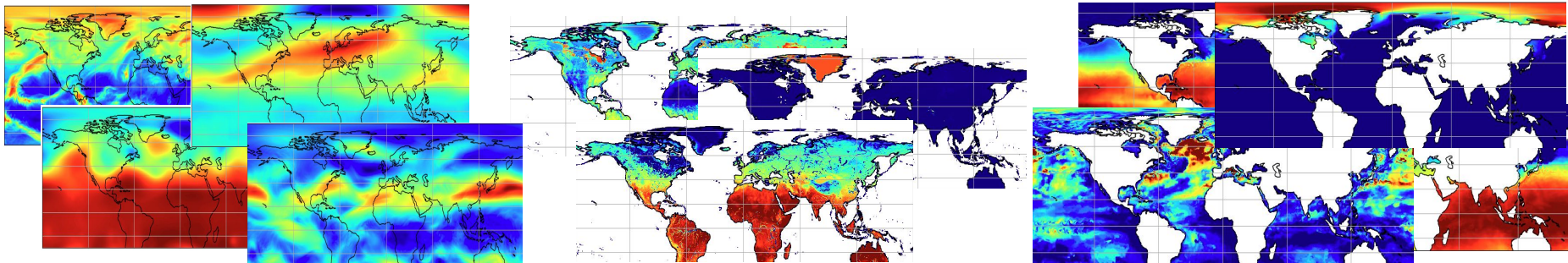
Predictors representing the initial state of the Earth system

Removed climatology

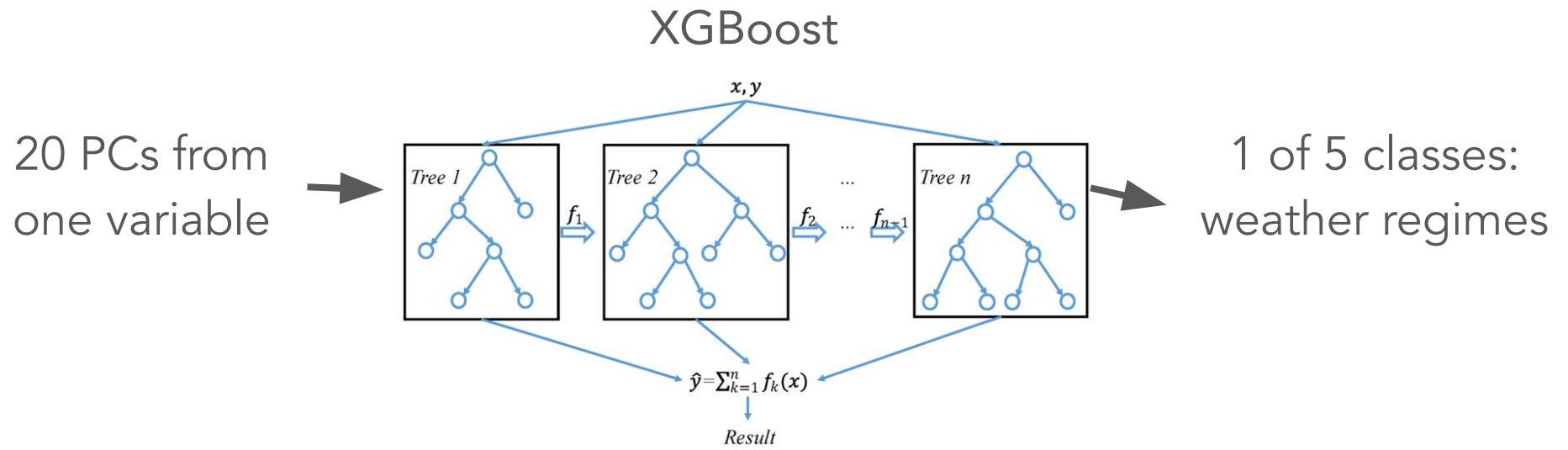
Removed trends

Trained different models with each variable
individually

Dimensionality reduction with PCA: For
each variable, we extracted the first 20 PCs



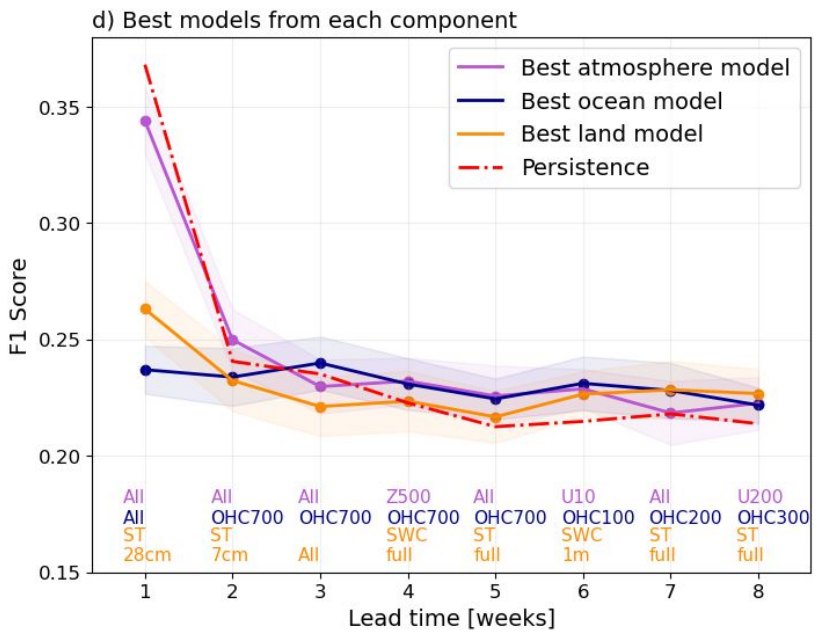
ML framework for each variable



Not looking for the best performance but for a fair comparison among models.

Friedman, J. H., 2001: Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
Molina, M. J., and Coauthors, 2023b: A review of recent and emerging machine learning applications for climate variability and weather phenomena. *Artificial Intelligence for the Earth Systems*, 1–46.609
Fatima, S., A. Hussain, S. B. Amir, S. H. Ahmed, and S. M. H. Aslam, 2023: Xgboost and random forest algorithms: An in depth analysis. *Pakistan Journal of Scientific Research*, 3 (1), 26–31.

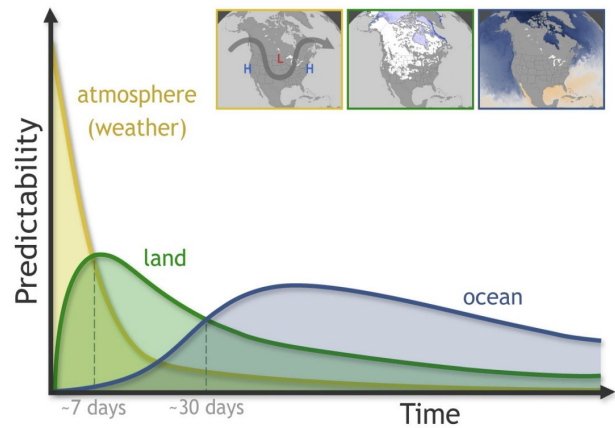
Models' comparison: Which component is more important?



Ocean OHC

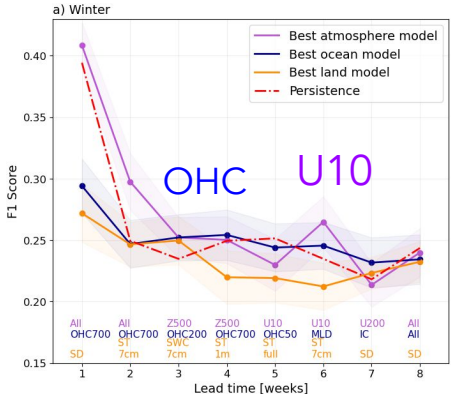
Land Integrated soil moisture and soil temperature

Atmosphere Combined model

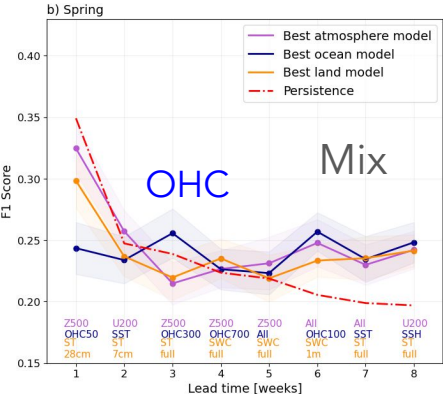


Seasonal differences

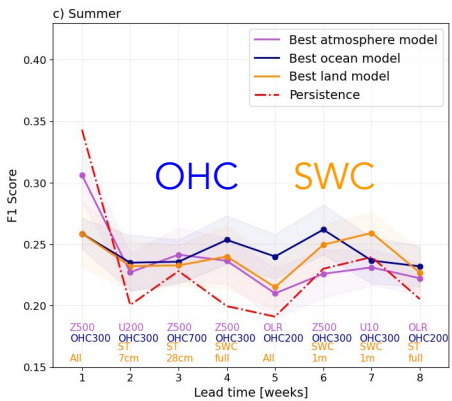
Winter



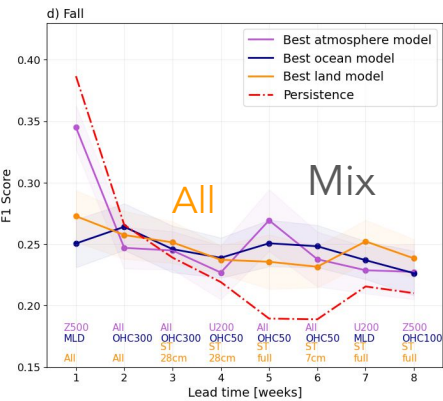
Spring



Summer



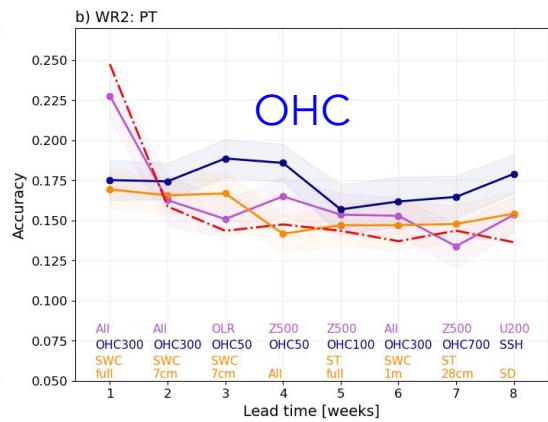
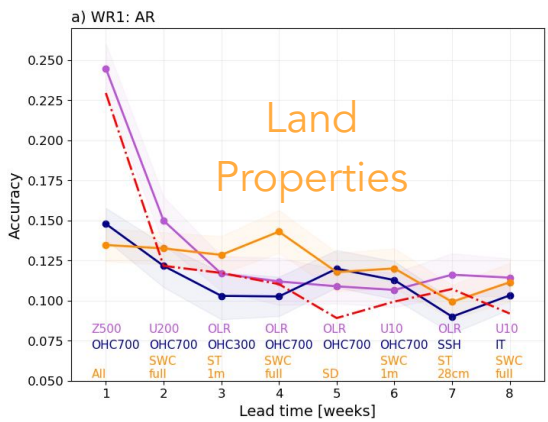
Fall



Differences depending on the WR of interest

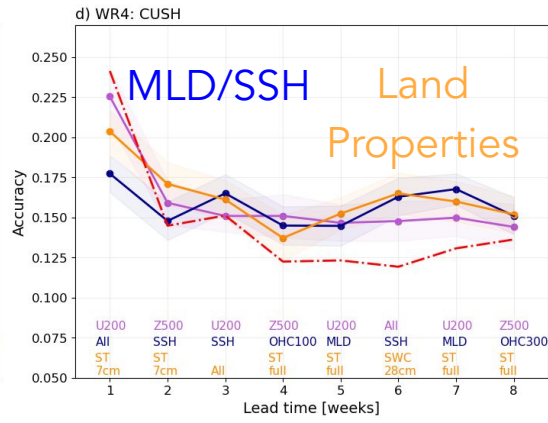
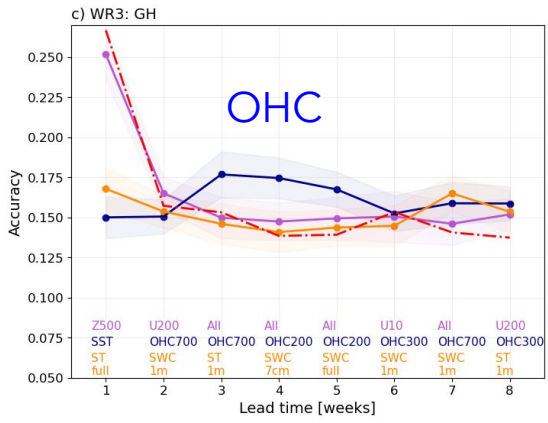
Alaskan Ridge

Pacific Trough



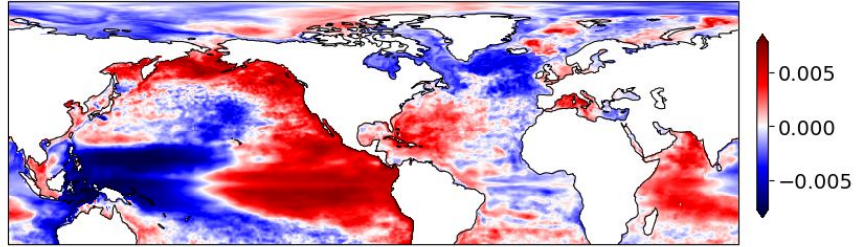
Greenland High

Central-US High



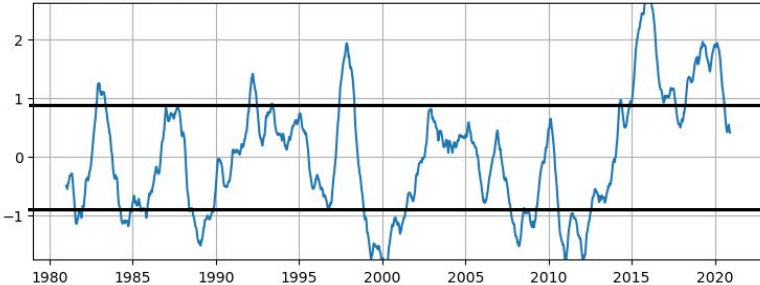
Analysis of the Principal Components

Variable: OHC300_SODA
PC: 0
Variance explained: 9.68%

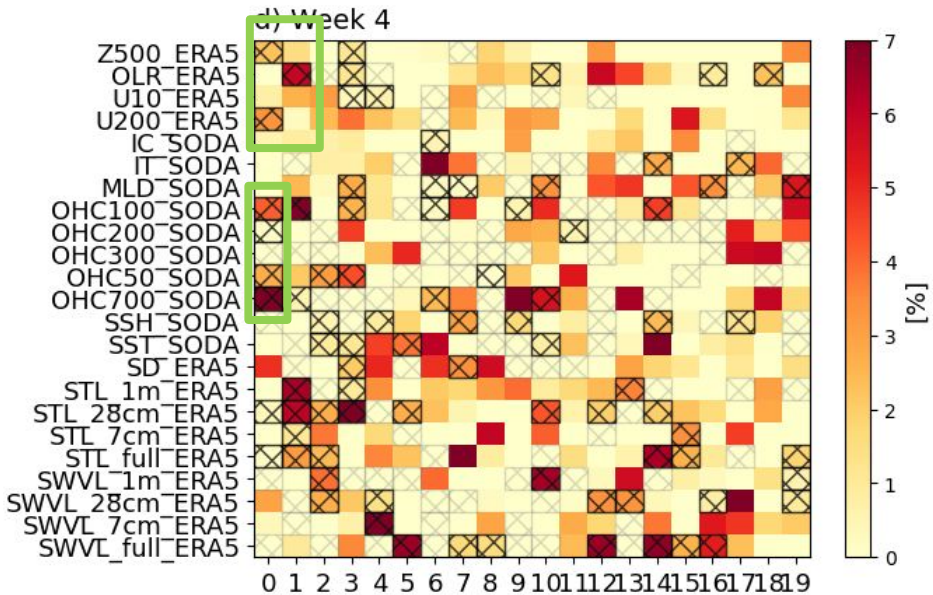


What spatial patterns are relevant from each variable?

Normalized time series



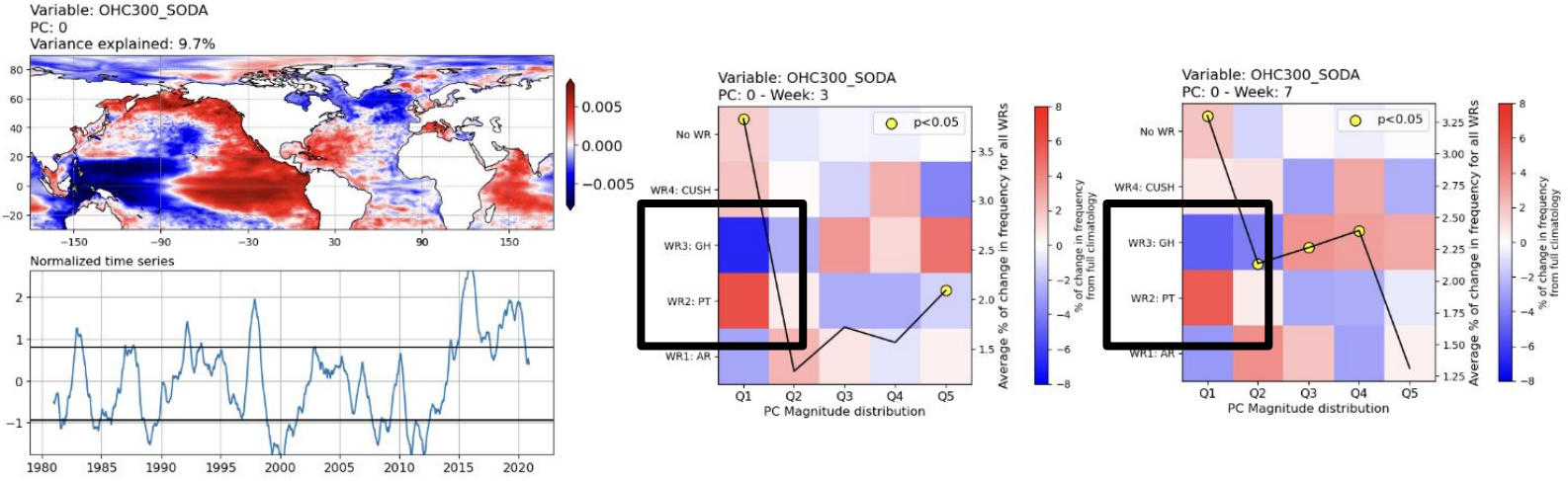
Which Principal Components are important?



2 Conditions (black hatching):

- The PC time series being anomalously high or low produced a statistically significant change in the distribution of future WRs.
- The magnitude of the PC time series is higher than average when correct predictions are made.

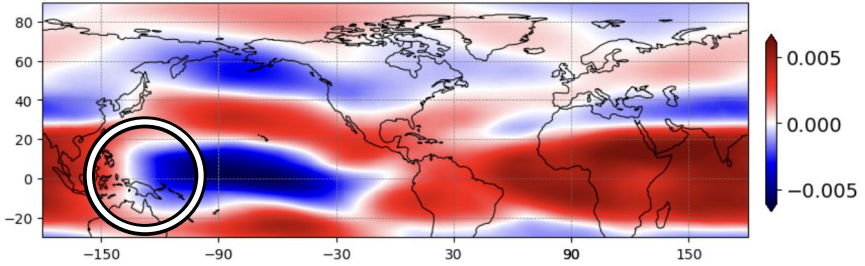
ENSO pattern



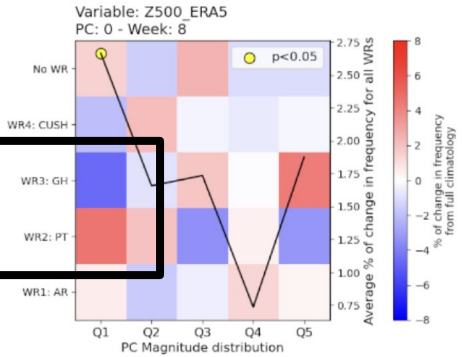
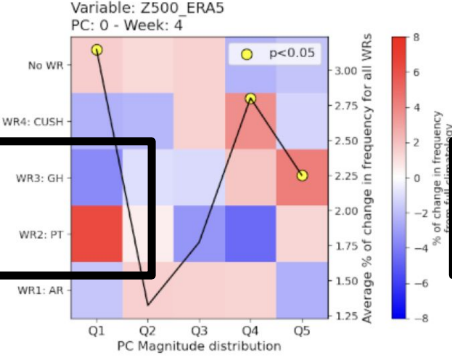
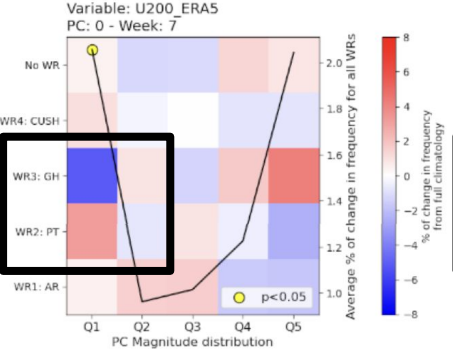
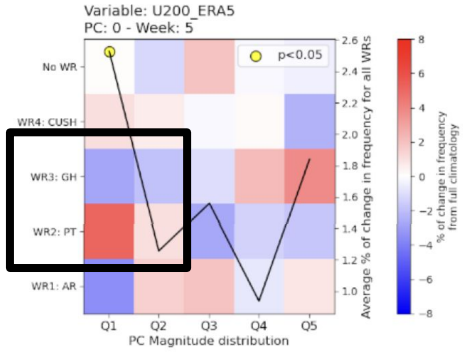
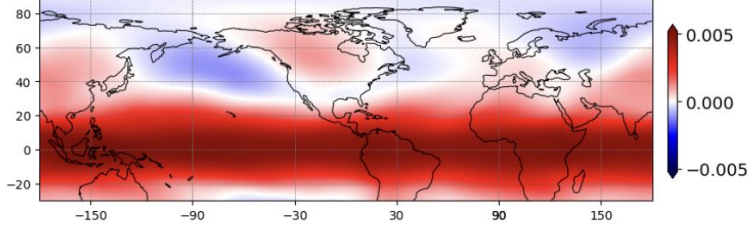
Same signal seen in OHC at different depths, sea-surface height, sea surface temperature.

Atmospheric Associated Pattern

Variable: U200_ERA5
 PC: 0
 Variance explained: 6.45%

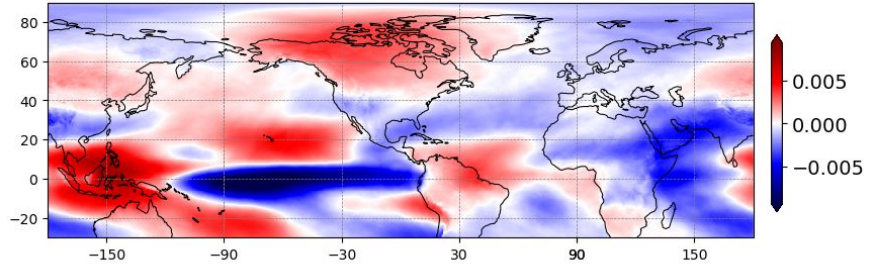


Variable: Z500_ERA5
 PC: 0
 Variance explained: 23.45%

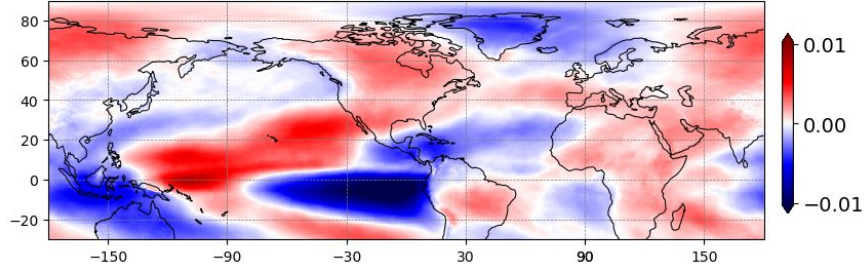


The effect changes depending on where convection is located

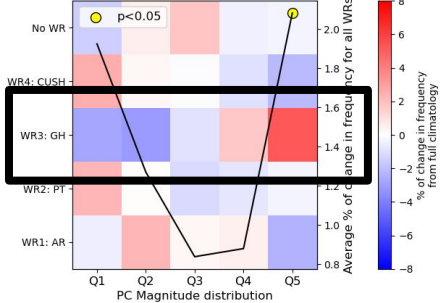
Variable: OLR_ERA5
PC: 1
Variance explained: 3.27%



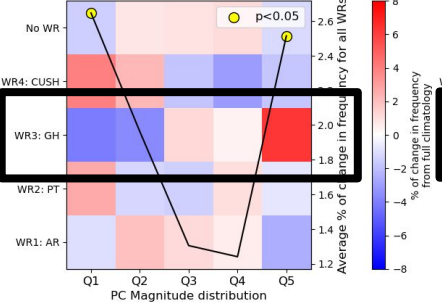
Variable: OLR_ERA5
PC: 3
Variance explained: 2.2%



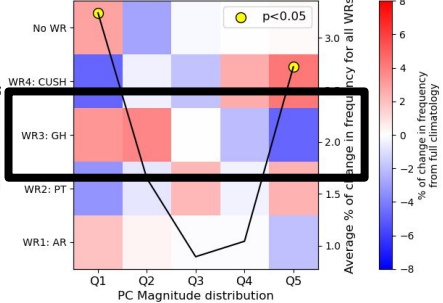
Variable: OLR_ERA5
PC: 1 - Week: 2



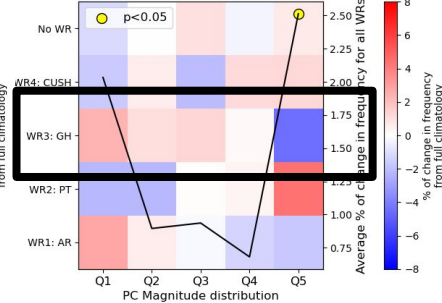
Variable: OLR_ERA5
PC: 1 - Week: 3



Variable: OLR_ERA5
PC: 3 - Week: 2

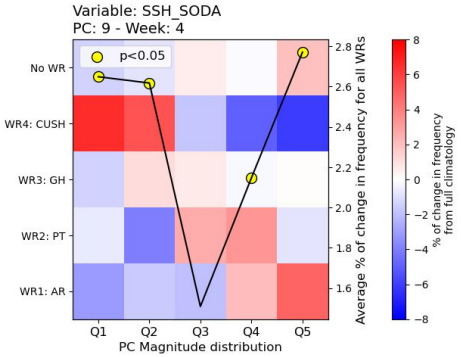
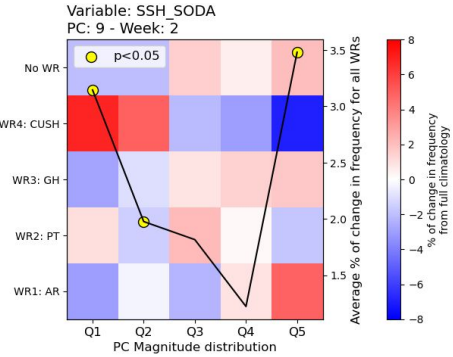
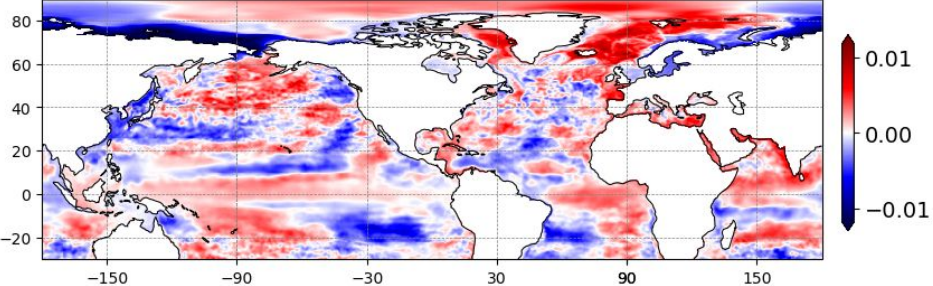


Variable: OLR_ERA5
PC: 3 - Week: 5

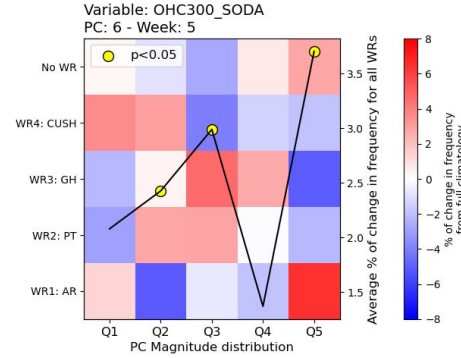
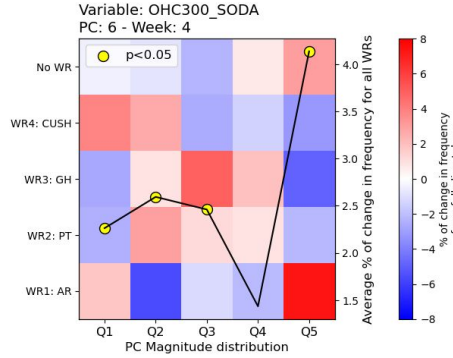
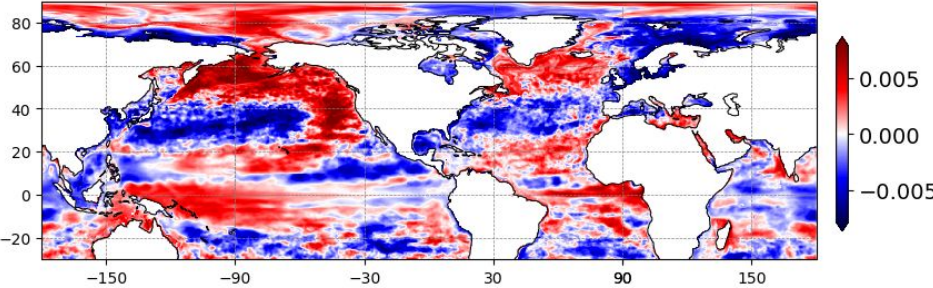


More from the ocean

Variable: SSH_SODA
 PC: 9
 Variance explained: 1.74%

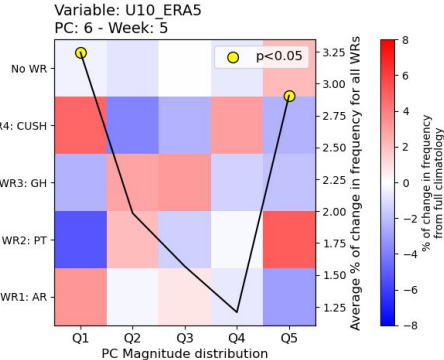
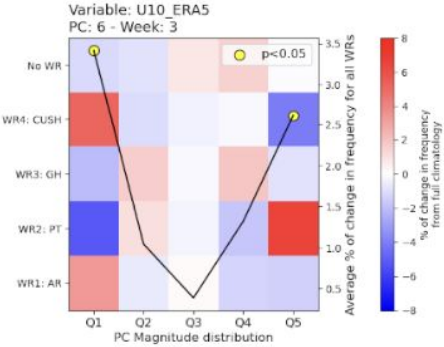
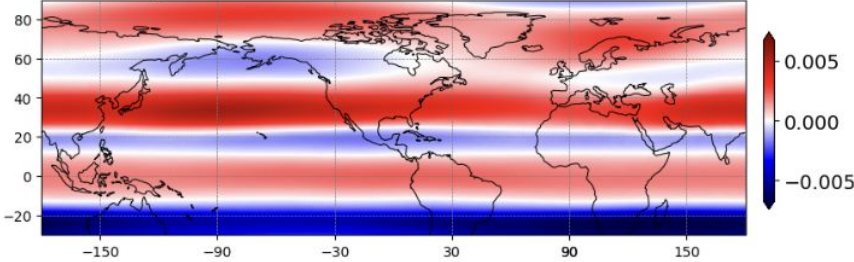


Variable: OHC300_SODA
 PC: 6
 Variance explained: 2.08%

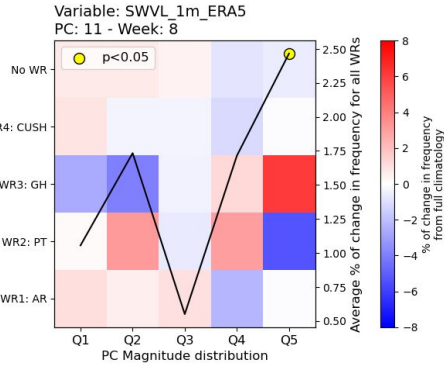
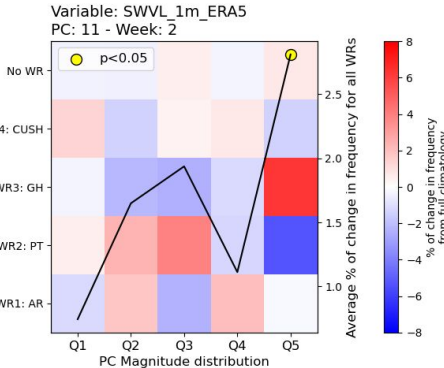
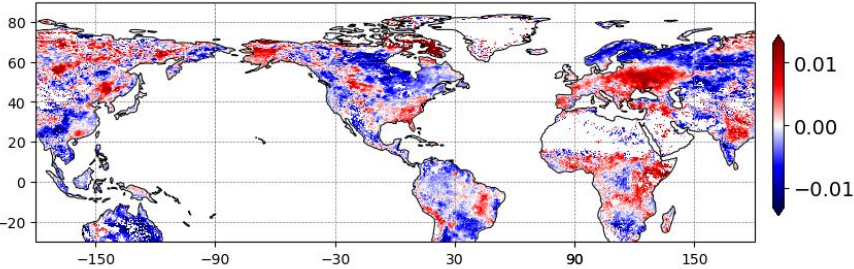


More from the atmosphere and land

Variable: U10_ERA5
 PC: 6
 Variance explained: 2.72%



Variable: SWVL_1m_ERA5
 PC: 11
 Variance explained: 1.44%

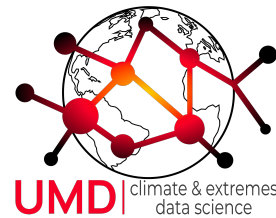


An Earth-System-Oriented View of the S2S Predictability of Weather Regimes using XGBoost

Jhayron S. Pérez-Carrasquilla^a and Maria J. Molina^{a,b}

^a *Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA*

^b *National Center for Atmospheric Research, Boulder, Colorado, USA*



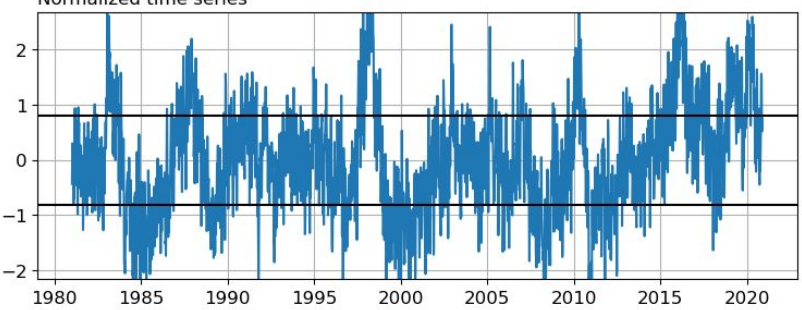
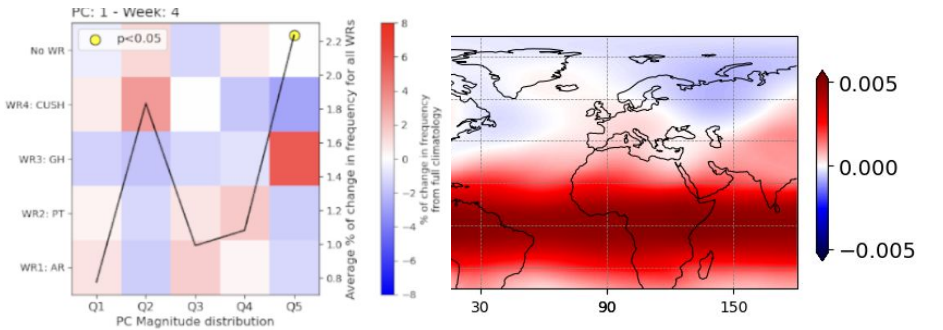
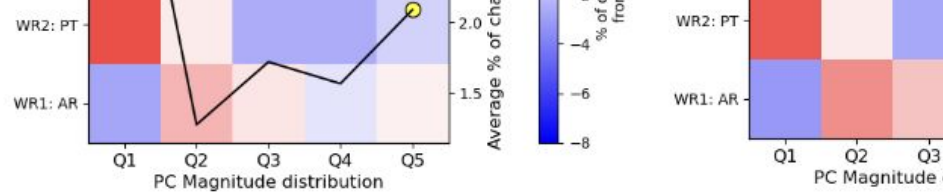
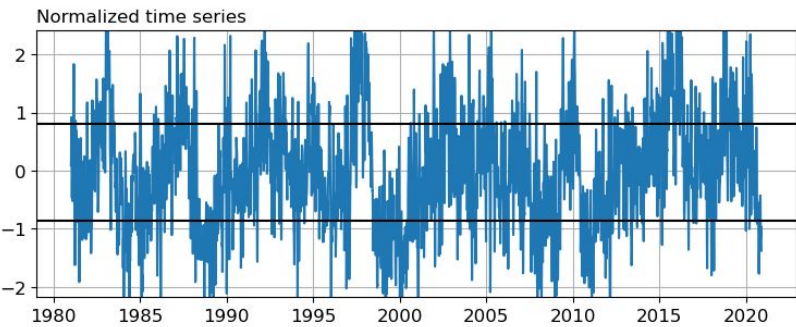
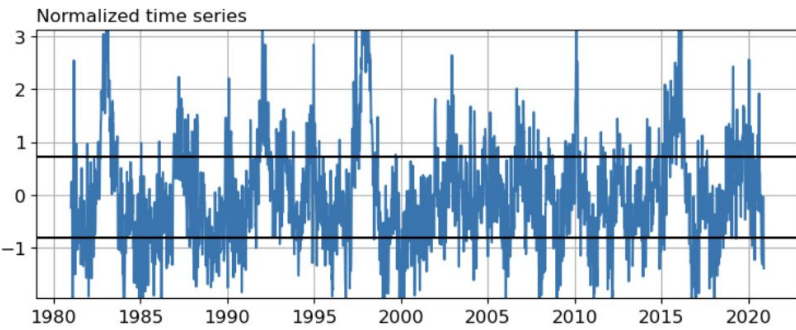
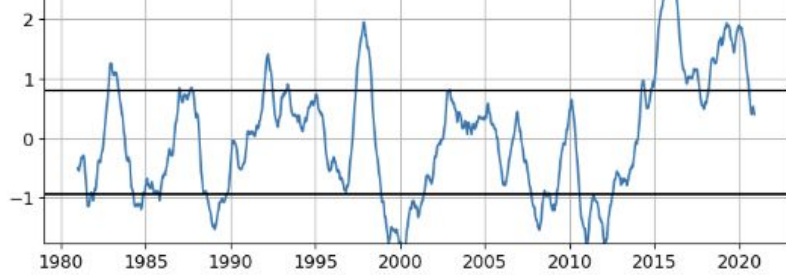
Summary

ML useful not only for prediction but for producing complementary scientific knowledge.

The ocean's role (OHC) appears to be the predominant source of information even from early stages. This is related to ENSO, which mainly affects Greenland High and Pacific Trough.

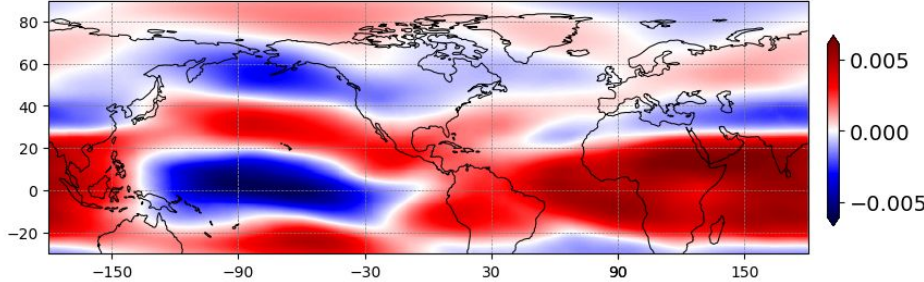
Other relevant processes identified:

- Initial state of Rossby/planetary waves
- Tropical convection: MJO, Indo-Pacific region
- Initial structure of the stratospheric jets.
- Structure of the walker circulation
- Ocean properties over the North Pacific Gyre and the Arctic Ocean.
- Soil moisture and temperature global distribution

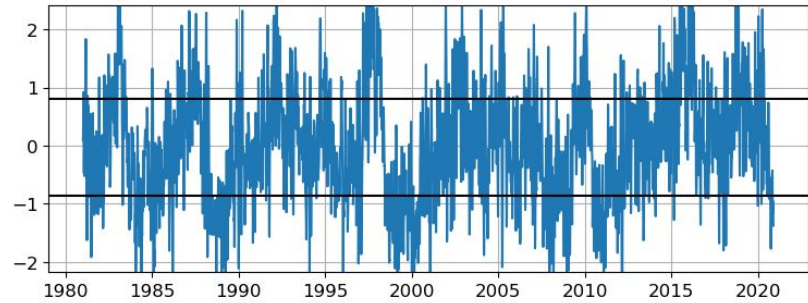


Atmospheric Associated Pattern

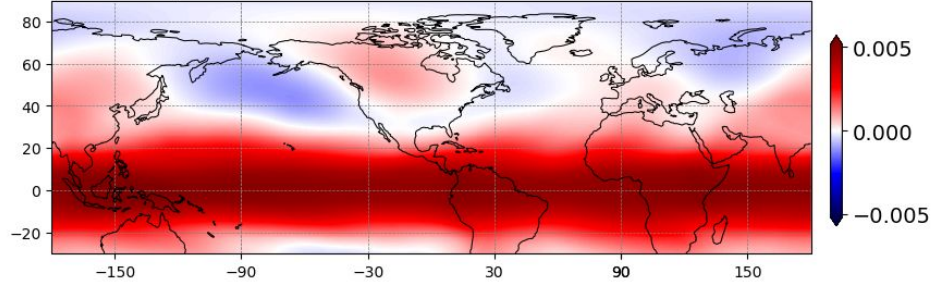
Variable: U200_ERA5
PC: 0
Variance explained: 6.45%



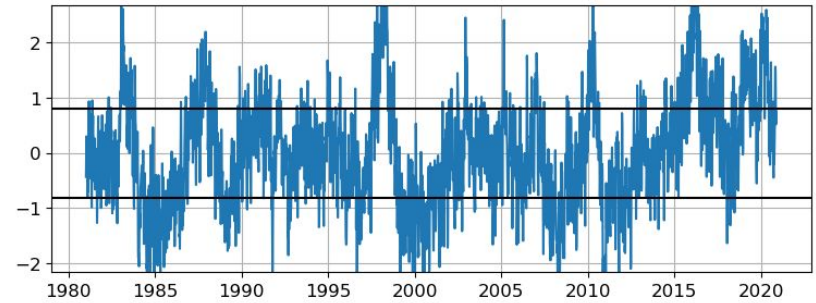
Normalized time series



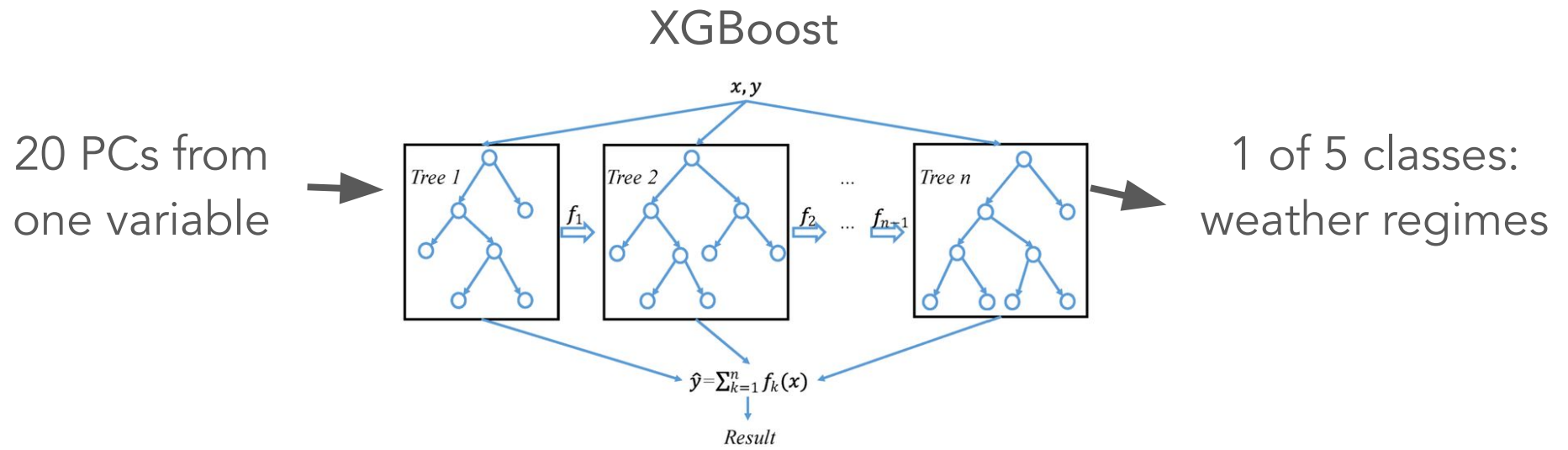
Variable: Z500_ERA5
PC: 0
Variance explained: 23.45%



Normalized time series



ML framework for each variable



Advantages: Widely used, similar performance to DL, low computational cost, fewer hyperparameters and lower sensitivity to them

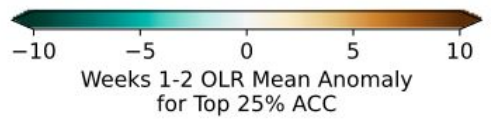
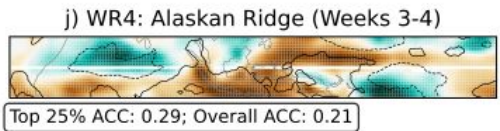
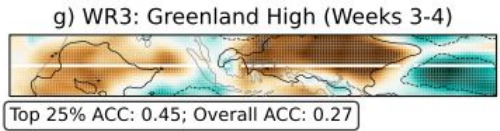
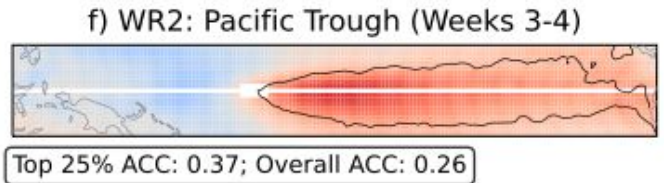
Friedman, J. H., 2001: Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
Molina, M. J., and Coauthors, 2023b: A review of recent and emerging machine learning applications for climate variability and weather phenomena. *Artificial Intelligence for the Earth Systems*, 1–46.609
Fatima, S., A. Hussain, S. B. Amir, S. H. Ahmed, and S. M. H. Aslam, 2023: Xgboost and random forest algorithms: An in depth analysis. *Pakistan Journal of Scientific Research*, 3 (1), 26–31.

Previous work on WRs predictability

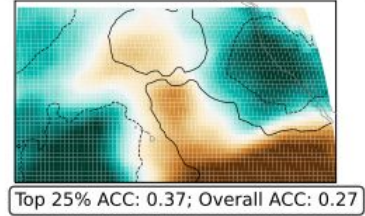
“Large-scale OLR and SST patterns upstream of North America across the subtropics and tropics can shed light on precursor mechanisms associated with weather regimes that can result in anomalous temperature, precipitation, and extremes”

“Skillful representation of upstream precursor patterns during weeks 1–2 within CESM2 can contribute to skillful prediction of weeks 3–4 weather regimes over North America ”

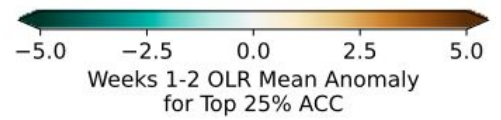
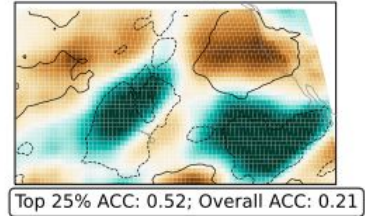
Previous work on WRs predictability



h) WR3: Greenland High (Weeks 3-4)

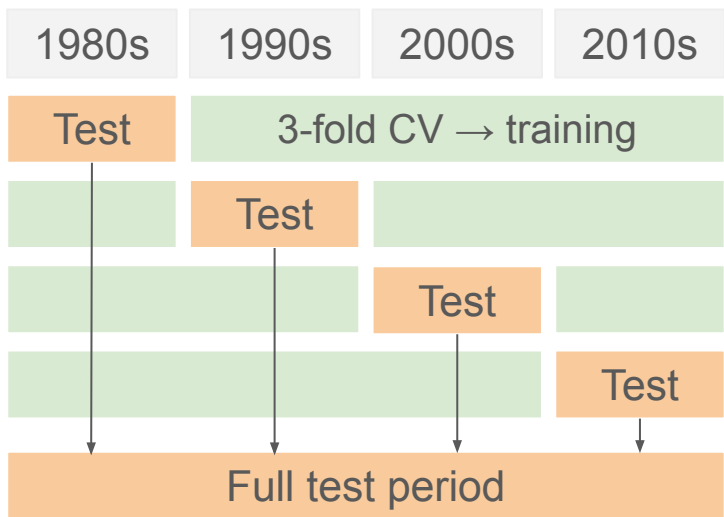


k) WR4: Alaskan Ridge (Weeks 3-4)



Bayesian hyperparameter optimization for xgboost and cross-testing/cross-validation

Aiming to find **robust** and **fair** models.



```
#### Tree specific hyperparameters ####
```

```
'max_depth': (3,12), #maximum depth of tree
```

```
'min_child_weight': (1,50), #minimum sum of instance weight needed in a child,  
#prevents the creation of too small leaves
```

```
'subsample': (0.1, 1), ## percentage of samples used for each tree construction
```

```
'colsample_bytree': (0.1, 1), ## percentage of features used for each tree construction.
```

```
'colsample_bylevel': (0.1, 1),## percentage of features used for each split/level.
```

```
#### Learning task-specific hyperparameters ####
```

```
'learning_rate': (0.01, 0.3), #step size shrinkage usage in updates
```

```
'gamma':(0, 3), #minimum loss reduction required to make a further partition on a leaf node of tree
```

```
'reg_lambda':(0, 10), #L2 regularization term on weights
```

```
'reg_alpha':(0, 10),#L1 regularization term on weights
```

```
#### General ####
```

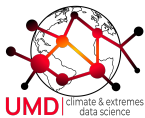
```
'class_weights':(0, 1), #use class weights True or False
```

176 models x 4 test-folds x 3 cv-folds
x ~100 trials =

211,200 trainings

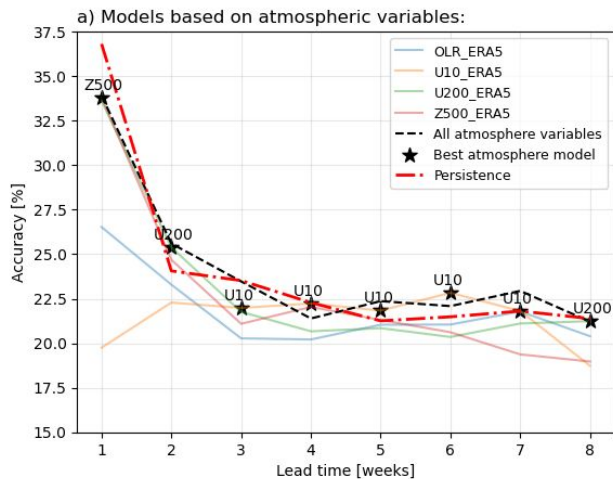


DEPARTMENT OF
ATMOSPHERIC &
OCEANIC SCIENCE

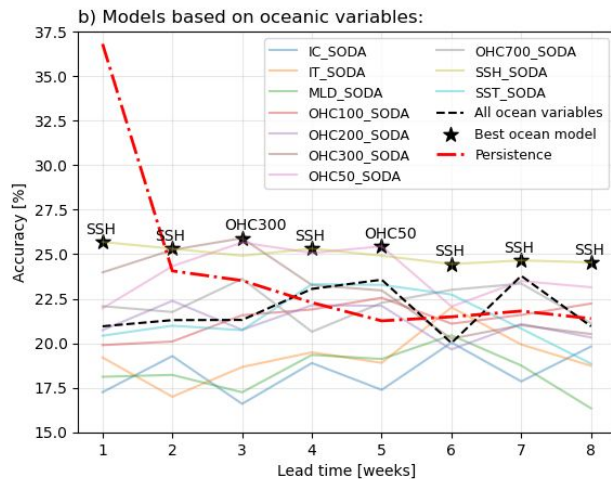


NCAR
CISL

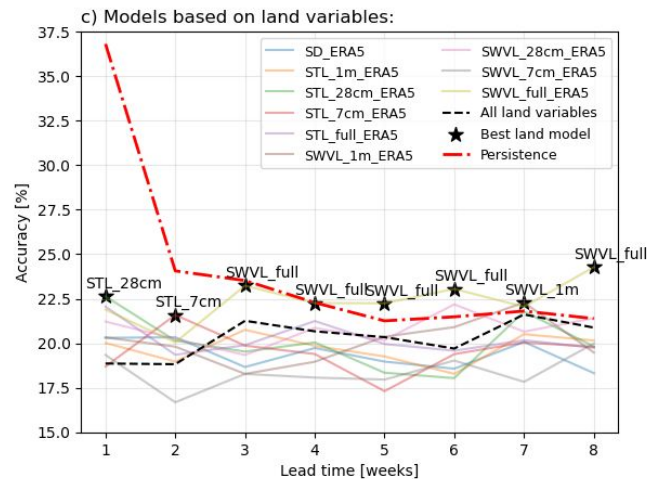
Overall results vs. persistence



Not much improvement over persistence, **stratospheric winds** performing the best.



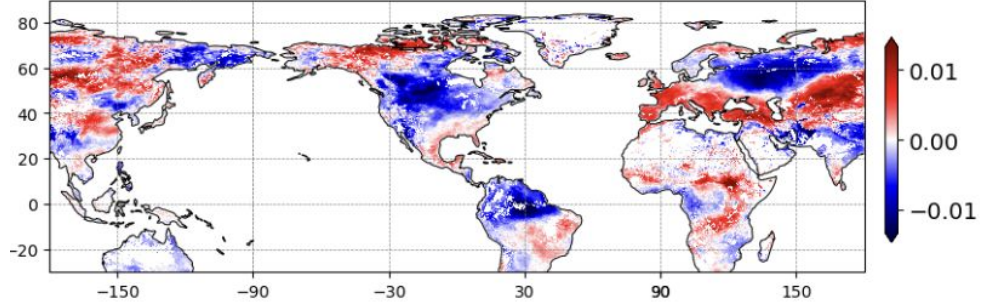
Better than persistence after week 2, **SSH and OHC** performing the best.



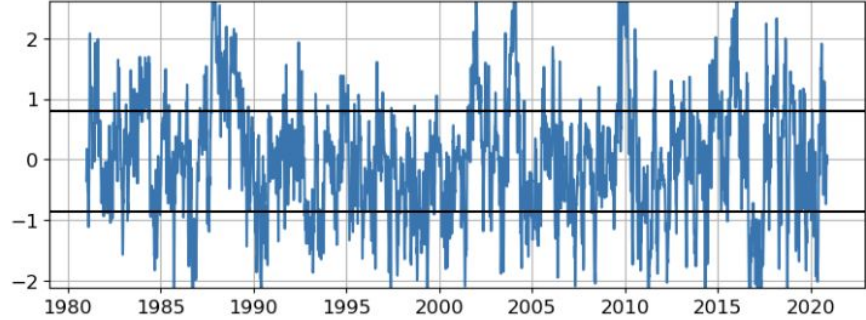
Small but increasing improvement over persistence after week 4, **soil moisture integrated down to 2.89m** performing the best.

Other precursors

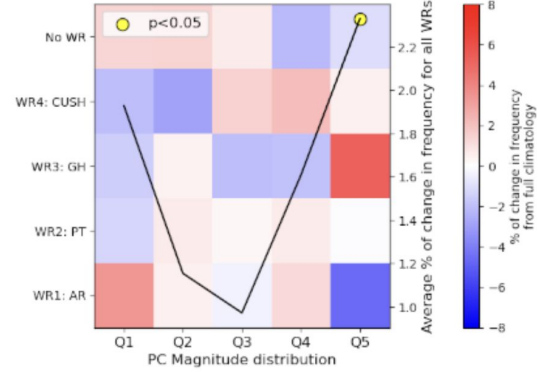
Variable: SWVL_7cm_ERA5
 PC: 7
 Variance explained: 1.39%



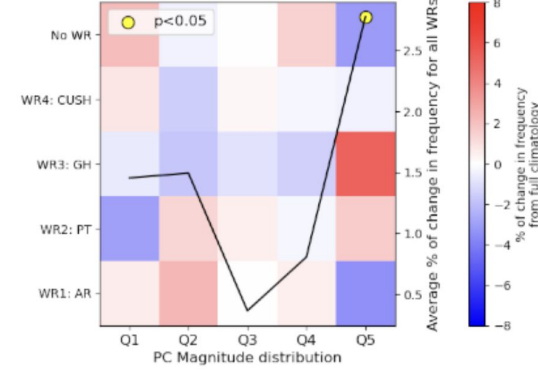
Normalized time series



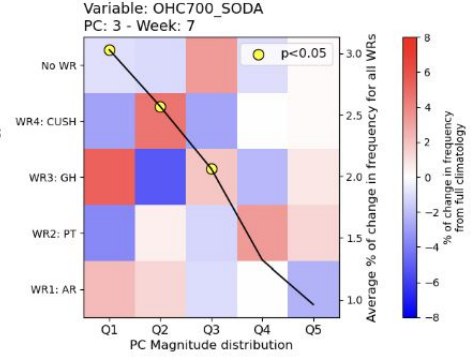
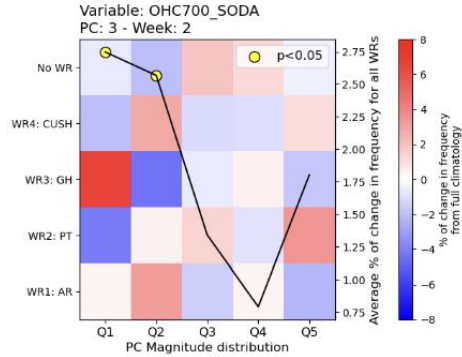
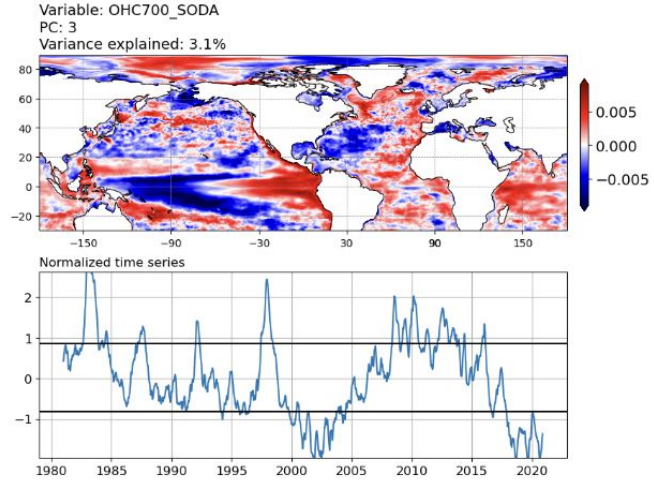
Variable: SWVL_7cm_ERA5
 PC: 7 - Week: 6



Variable: SWVL_7cm_ERA5
 PC: 7 - Week: 8



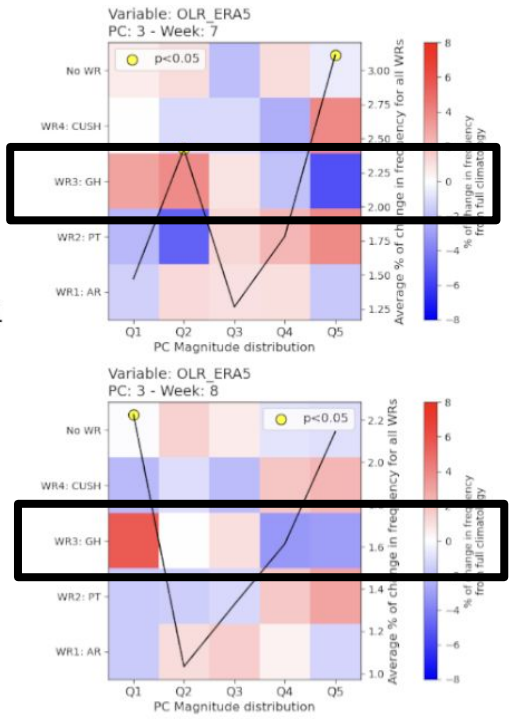
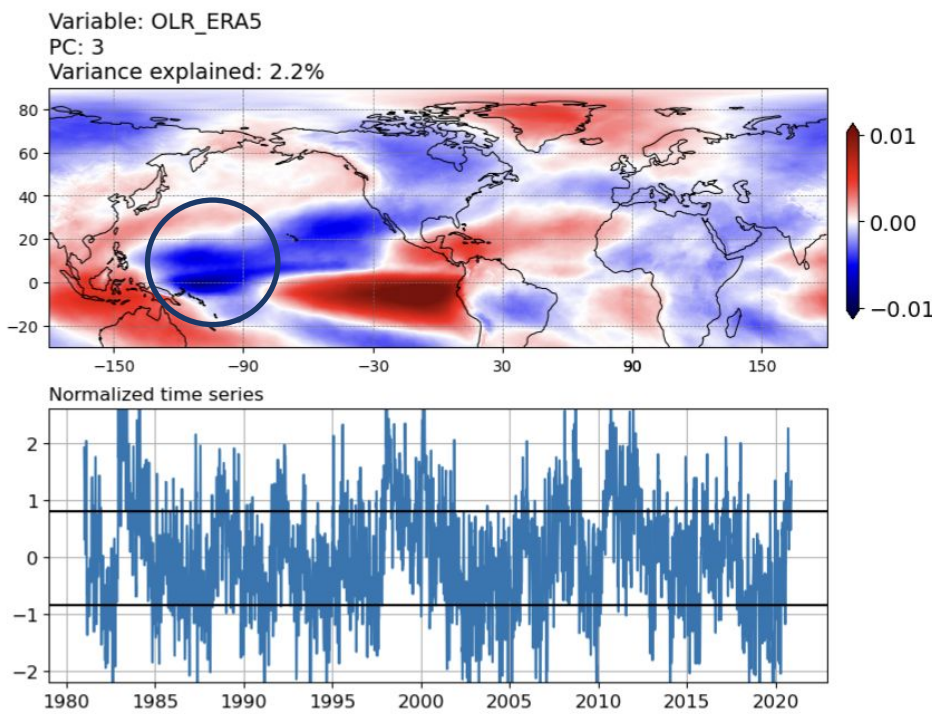
Oceanic predictors: Tropics



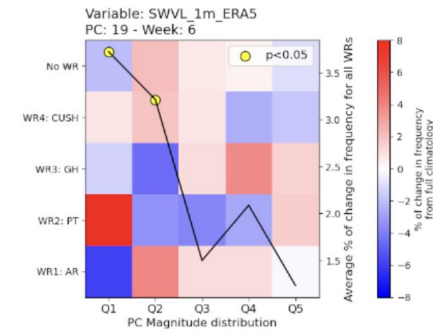
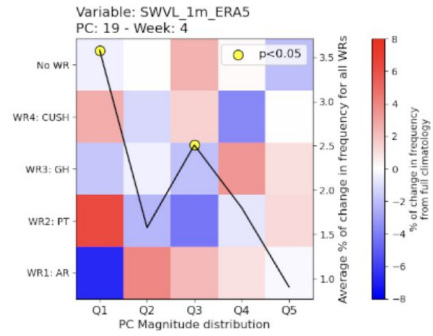
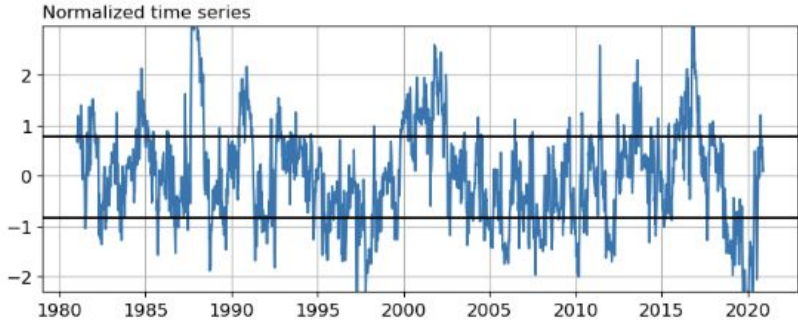
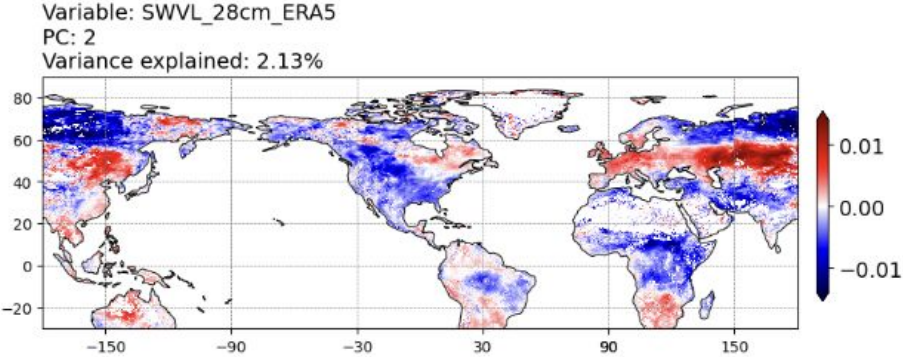
Same signal seen in OHC at different depths and sea-surface height.

The effect changes depending on where convection is located

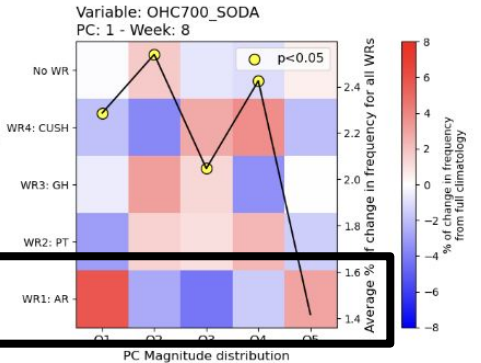
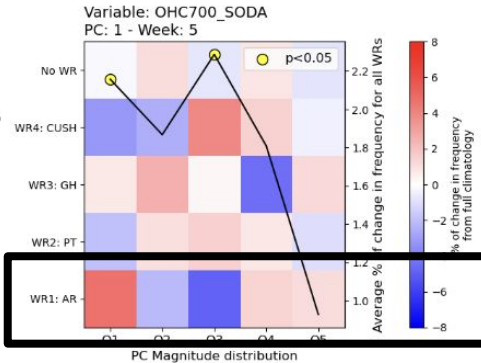
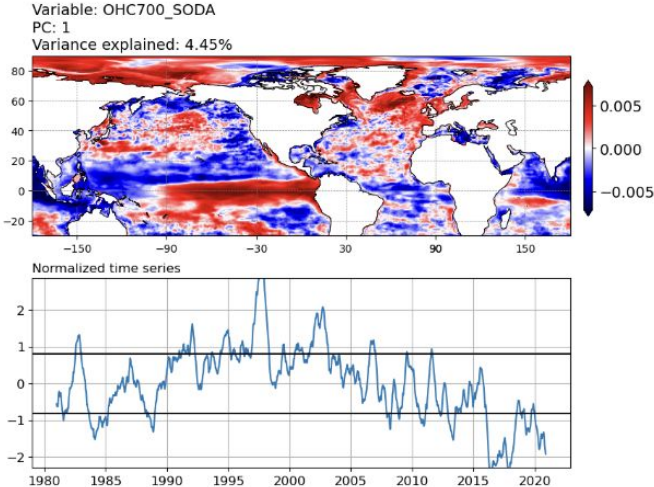
Predictability may "last" more if convection is more spread over the Pacific.



Other precursors



Oceanic predictors: ENSO + Arctic Ocean



Oceanic predictors: Indian Ocean

