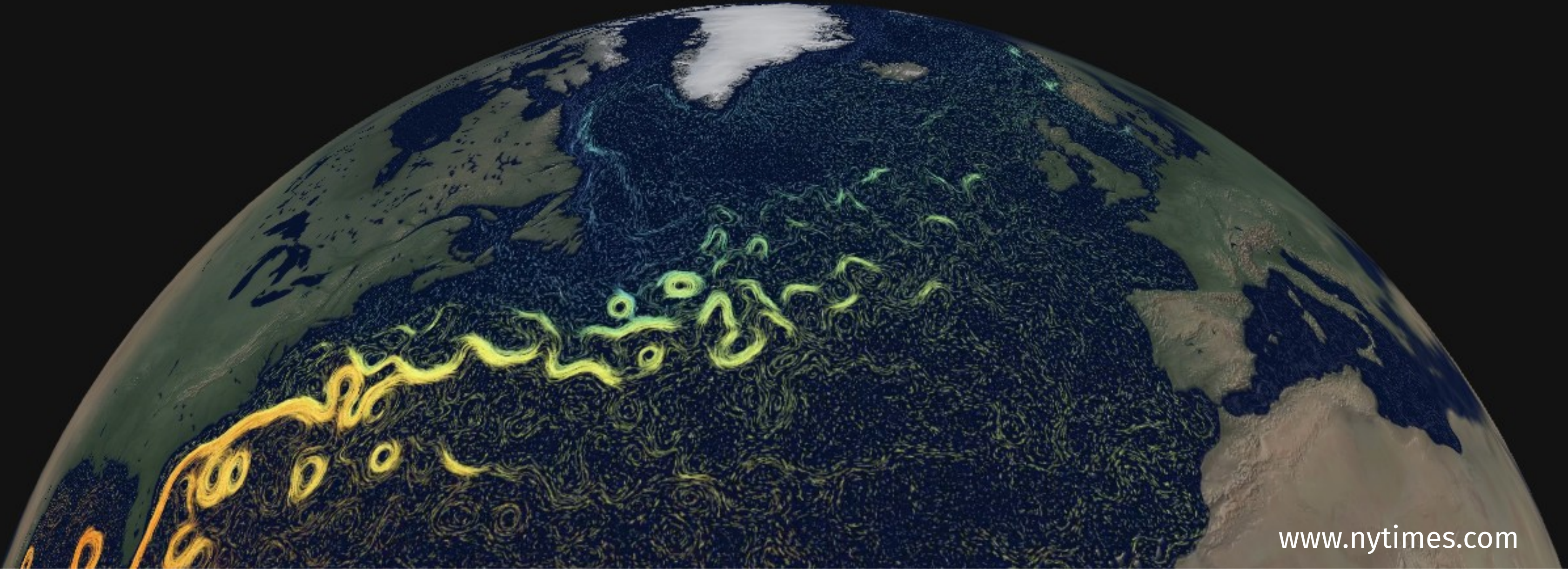
Uncertainties in observation-based Early Warning Signals of the AMOC

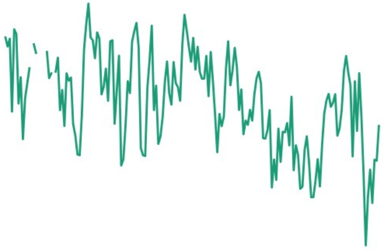
Maya Ben-Yami[1,2], Vanessa Skiba[2], Sebastian Bathiany[1,2], Niklas Boers[1,2,3]
maya.ben-yami@tum.de



[1] Earth System Modelling, Technical University of Munich, Munich, Germany
[2] Potsdam Institute for Climate Impact Research, Potsdam, Germany
[3] Department of Mathematics and Global Systems Institute, University of Exeter, Exeter, UK

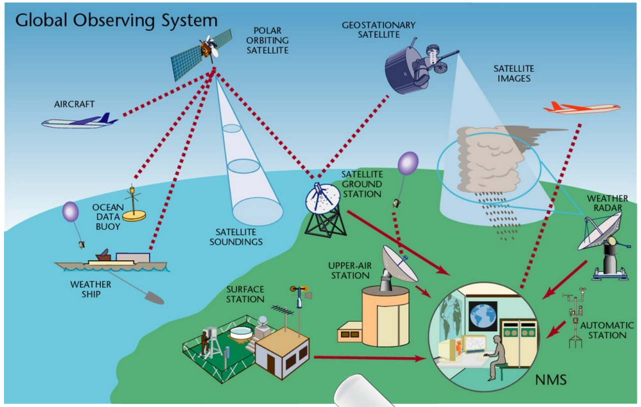
Expanded presentation with more text

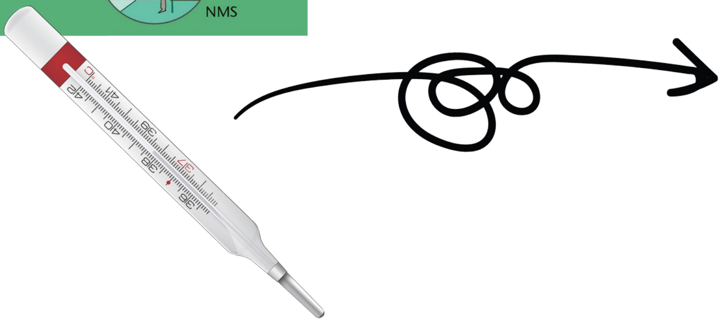
Dataset

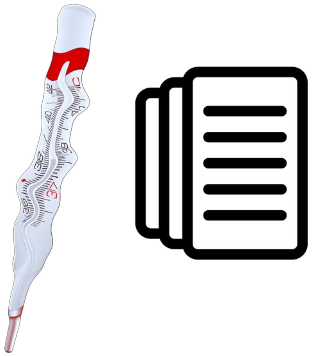


Data

Higher order statistic







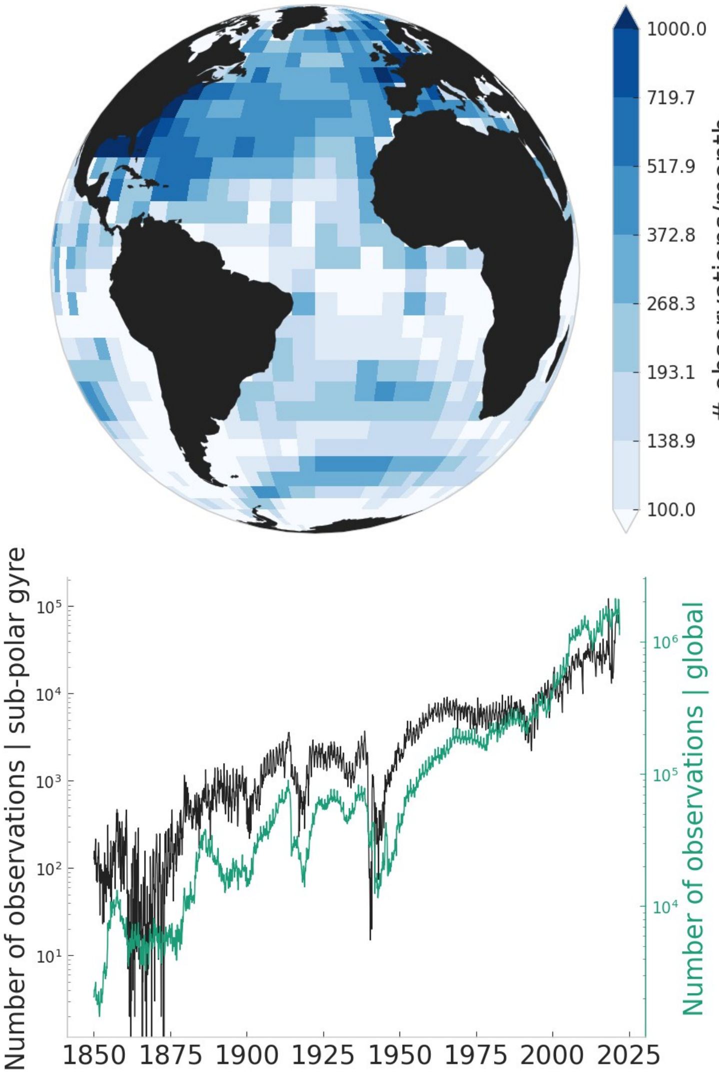
- We often use observational datasets to calculate higher order statistics
- But we should remember that the datasets are not raw measurement data, but have gone through a variety of processing and infilling steps
- These processing and infilling methods are different for each dataset
- The question is, could these processing methods be causing **false signals in the higher order statistics?**

AMOC

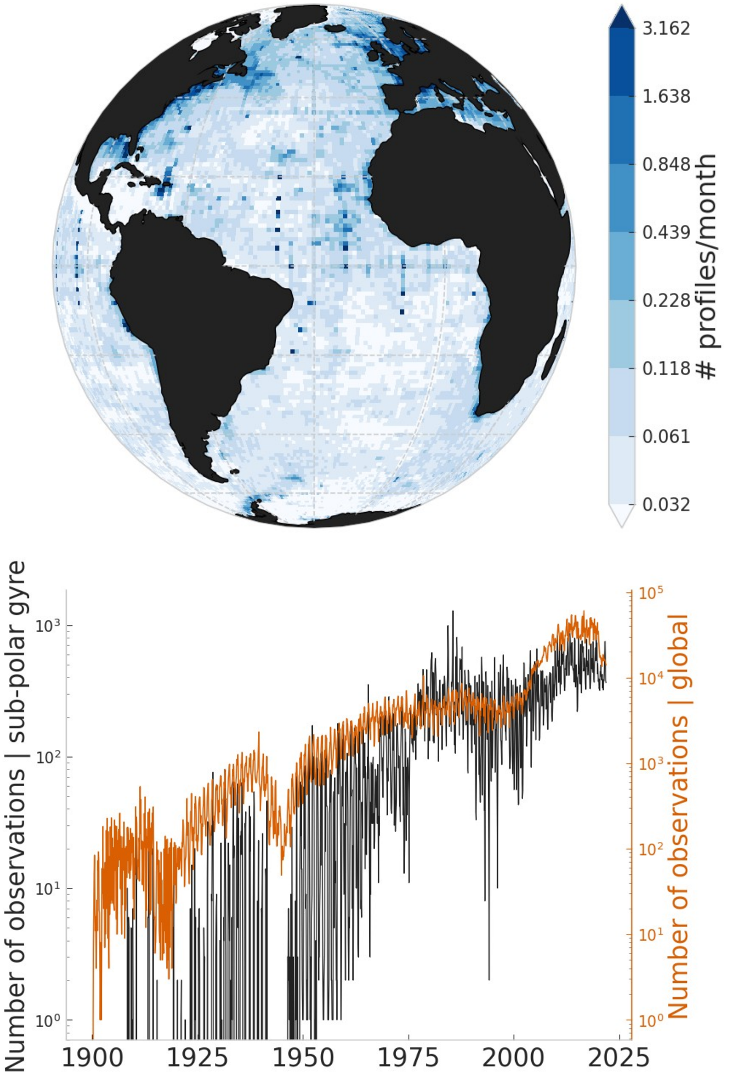
For the AMOC we use SST and salinity observational datasets to calculate early-warning signals (EWS).
→ we calculate EWS for both the average value in specific regions like the SPG and the time series at each grid cell

Both SST and salinity datasets are based on an uneven distribution of observations which increase exponentially from 1850 to today.

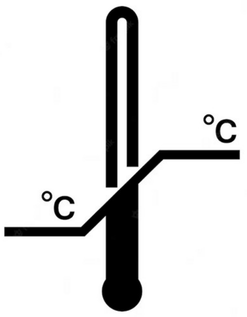
SST - HadCRUT5/HadSST4

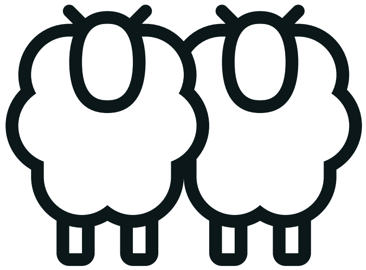


Salinity – EN4.2.2



How do we estimate the dataset effects on higher order statistics?





2 Use **uncertainty estimates** that come with the datasets

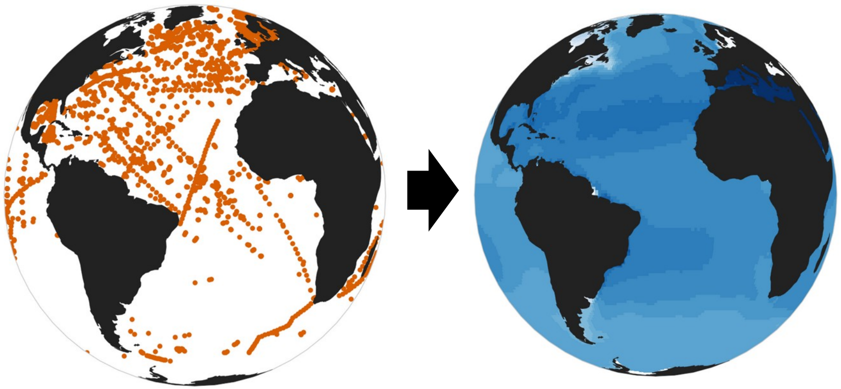
1 Use **surrogates** to estimate the significance of trends

Incorporate the effects of **data processing**

1 EN4.2.2 salinity dataset:

The infilled global analysis dataset is produced from scattered profiles by:

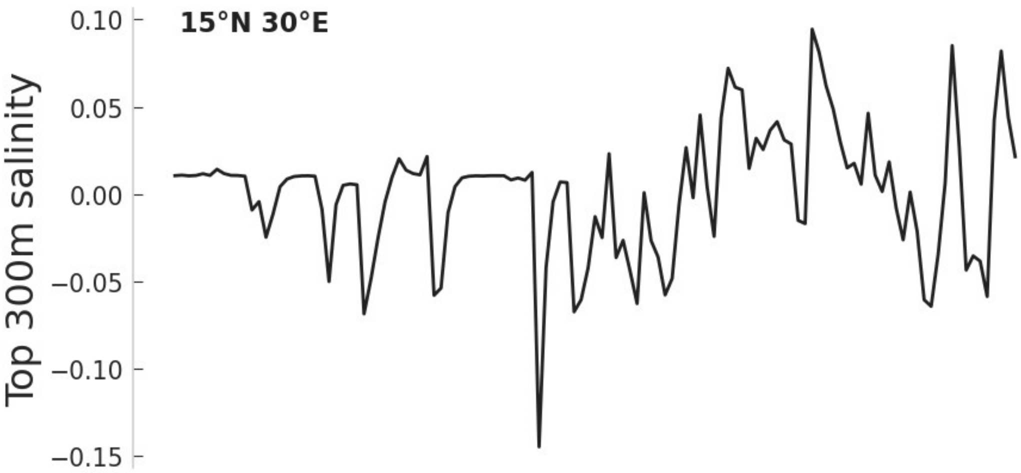
- A regularized least squares fit to the observations
- Using a persistence base forecast as an a-priori → this forecast (equation below) includes the climatology plus the previous months anomaly times a decay term



$$\mathbf{x}_i^f = \mathbf{x}_i^c + \alpha \underbrace{\left(\mathbf{x}_{i-1}^a - \mathbf{x}_{i-1}^c\right)}_{\text{Previous month anomaly}}$$

Climatology

0.9

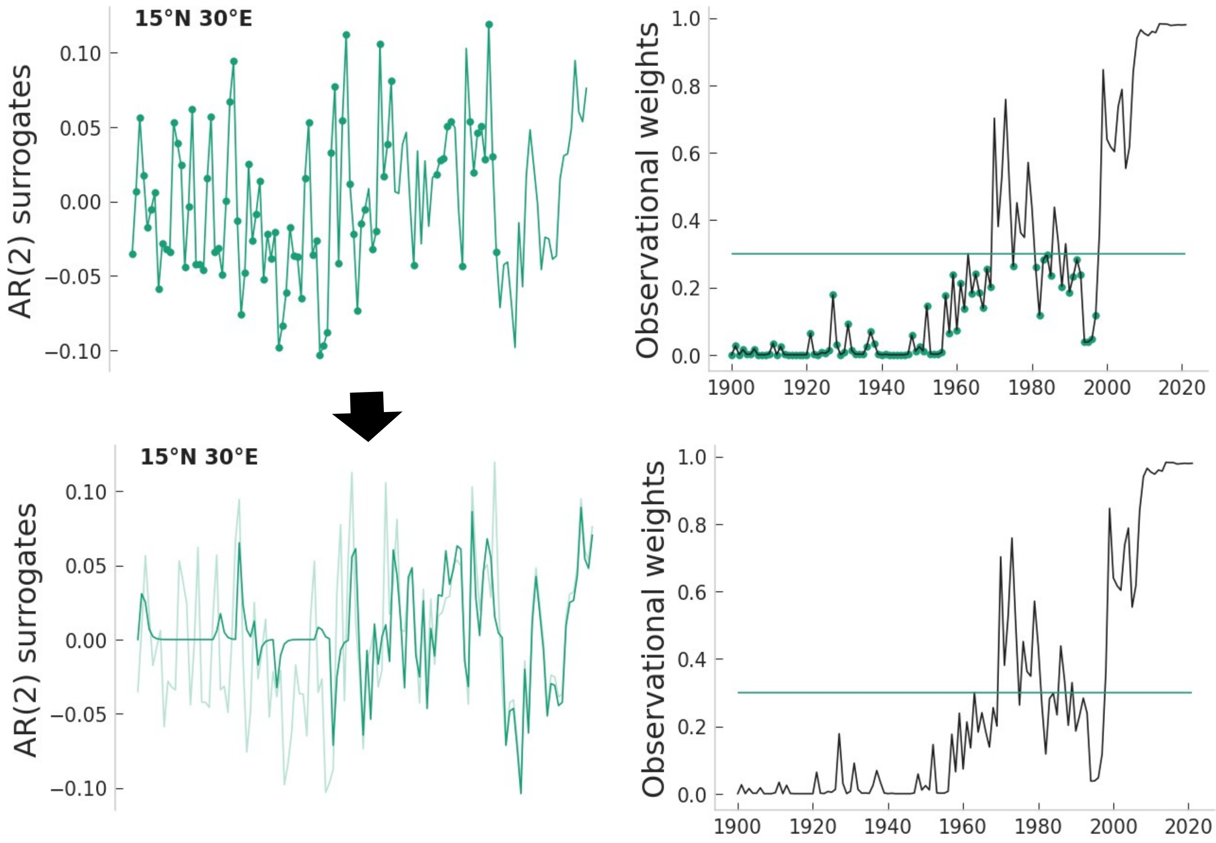


Using the persistence-based forecast as an a-priori results in time series that look like this: at earlier times it is **a flat line with spikes**.

Because there are few observations in the early 20th century, the flat line is the climatology, and the spikes happen when there is an observation at that location. After the value jumps to match the observation it relaxes back to the climatology with a timescale given by the equation on the previous slide.

Modification of EN4 surrogates:

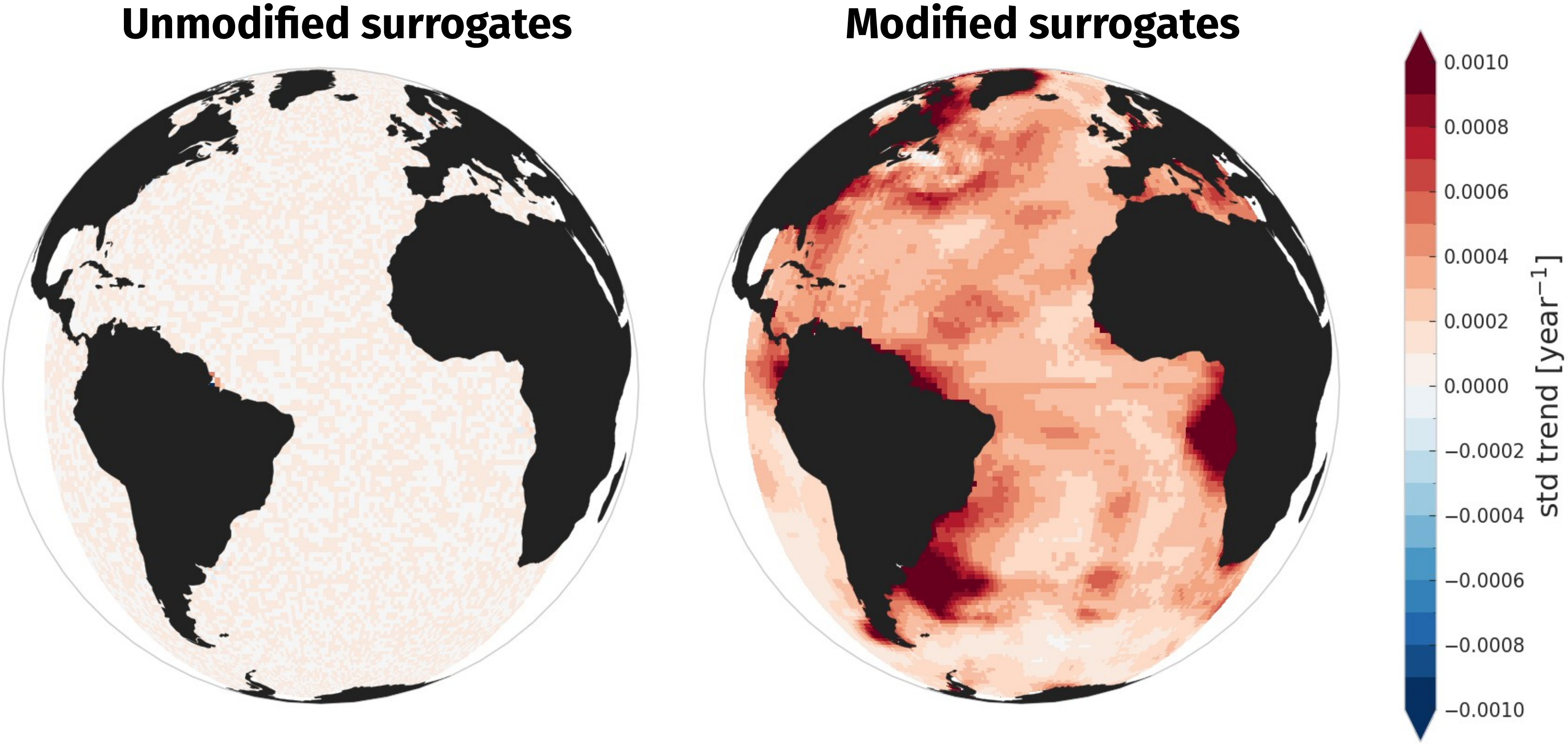
We use the observational weights provided with the dataset to modify the surrogates to match the data processing method.



We replace all data points that are below some observational weight **w** with the persistence based forecast, which results in a modified surrogates whose statistical properties have been affected in a similar way to the data in the EN4 dataset.

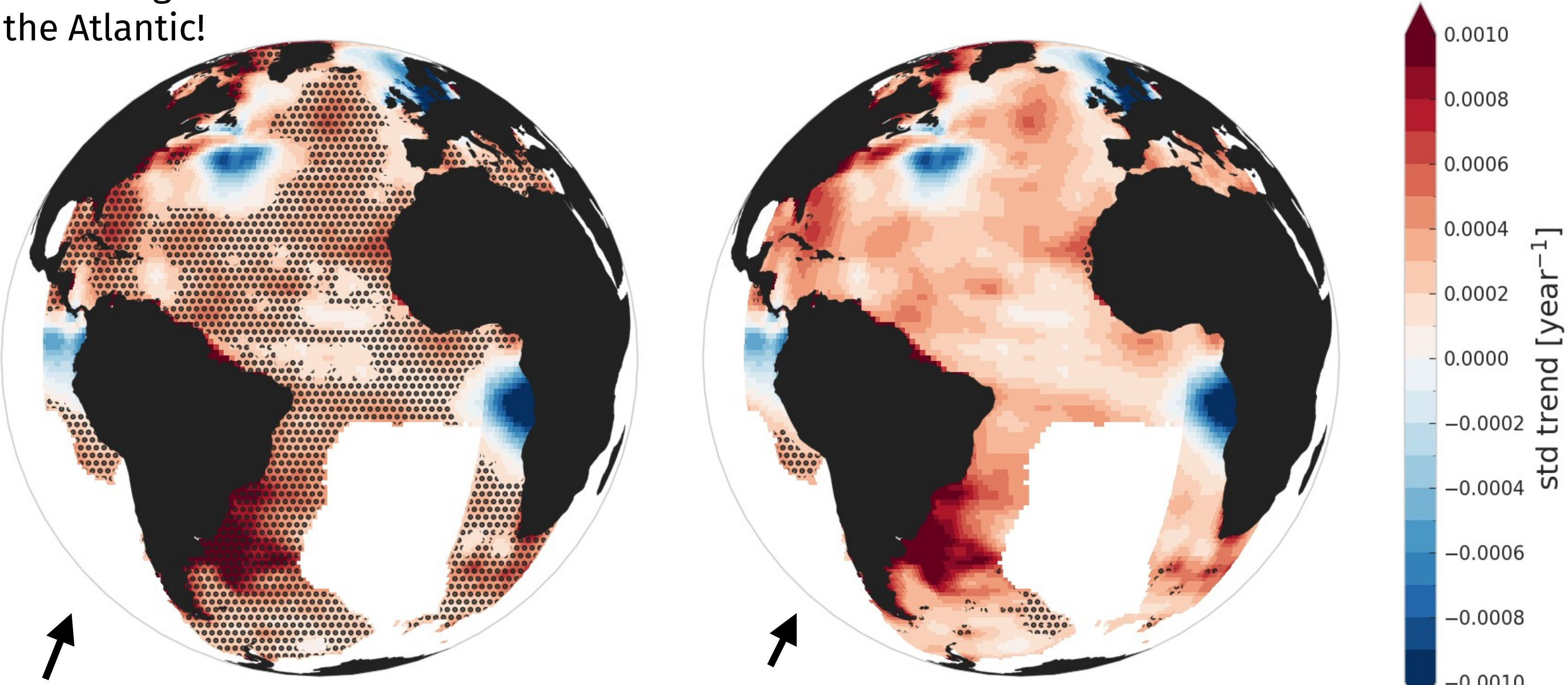
Mean variance trend of 1000 surrogates

The modification results in a false increase in variance!



Analysis data variance trend

We use the surrogates to calculate regions where the trend in the variance in the actual dataset is significant at a 95 percentile level. These are shown as black stippling.
→ modifying the surrogates reveals that in the EN4 dataset the increase in variance is not significant in the Atlantic!



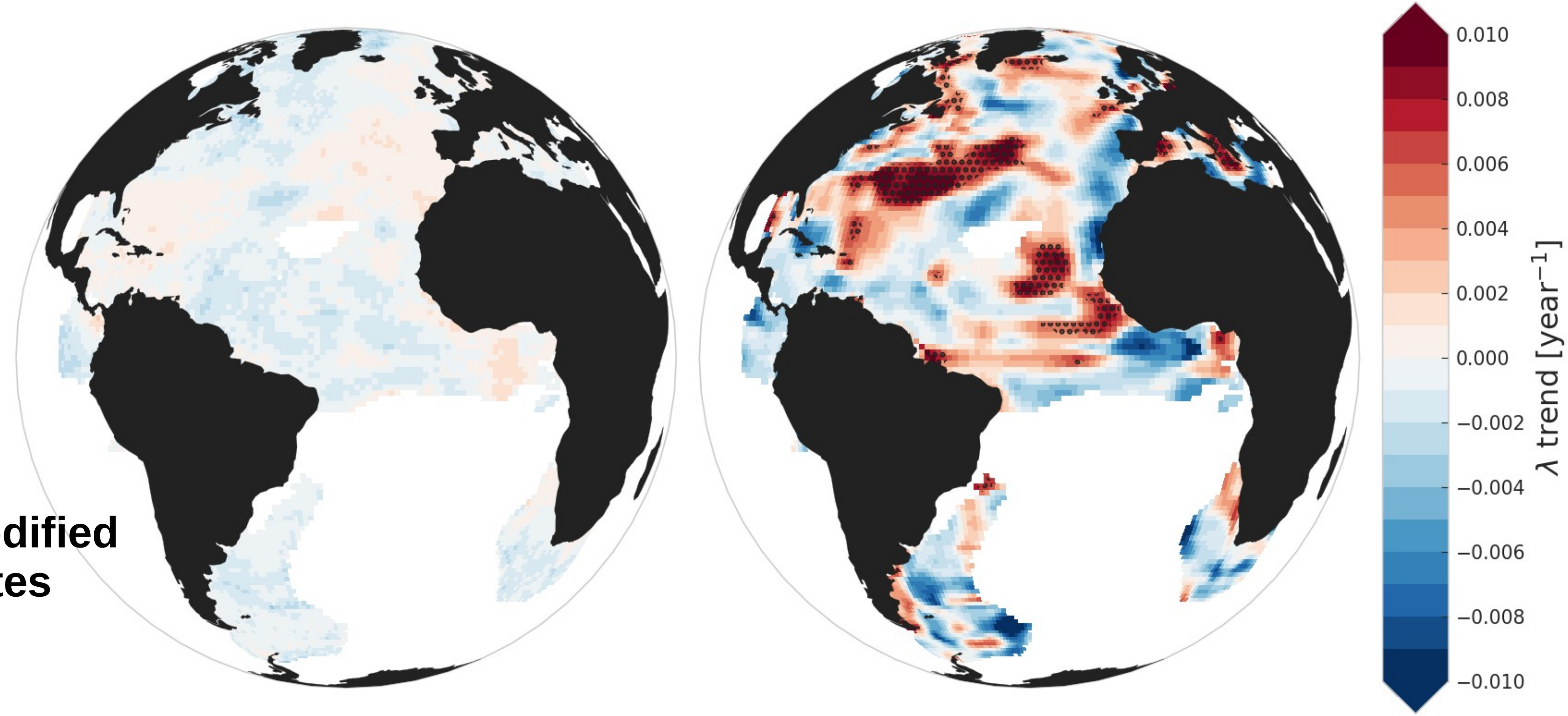
p<0.05 significance
unmodified surrogates

p<0.05 significance
modified surrogates

Restoring rate (λ)

For the restoring rate the situation is more complex – even though there are false increases in some regions, these are not strong enough to affect the significance of the increase in the dataset.

Dataset
significant regions stippled



Mean of modified surrogates

1 Surrogates

Overall, we use two different types of surrogates to calculate significant regions for three different datasets

Fourier surrogates

FFT → * random phases → IFFT

By the Wiener-Khinchin theorem, **preserves the variance and autocorrelation function**

Use when:
1. No gaps in time series
2. The autocorrelation function of the full time series is realistic

AR(2) surrogates

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \epsilon_t$$

Get the **a_1, a_2 that give the closest AR(1)** to the data on a monthly and yearly resolution for a part of the time series

Use when:
1. Gaps in time series
2. Want to capture the autocorrelation of only part of the time series (e.g. last 40 years)

1 Surrogates

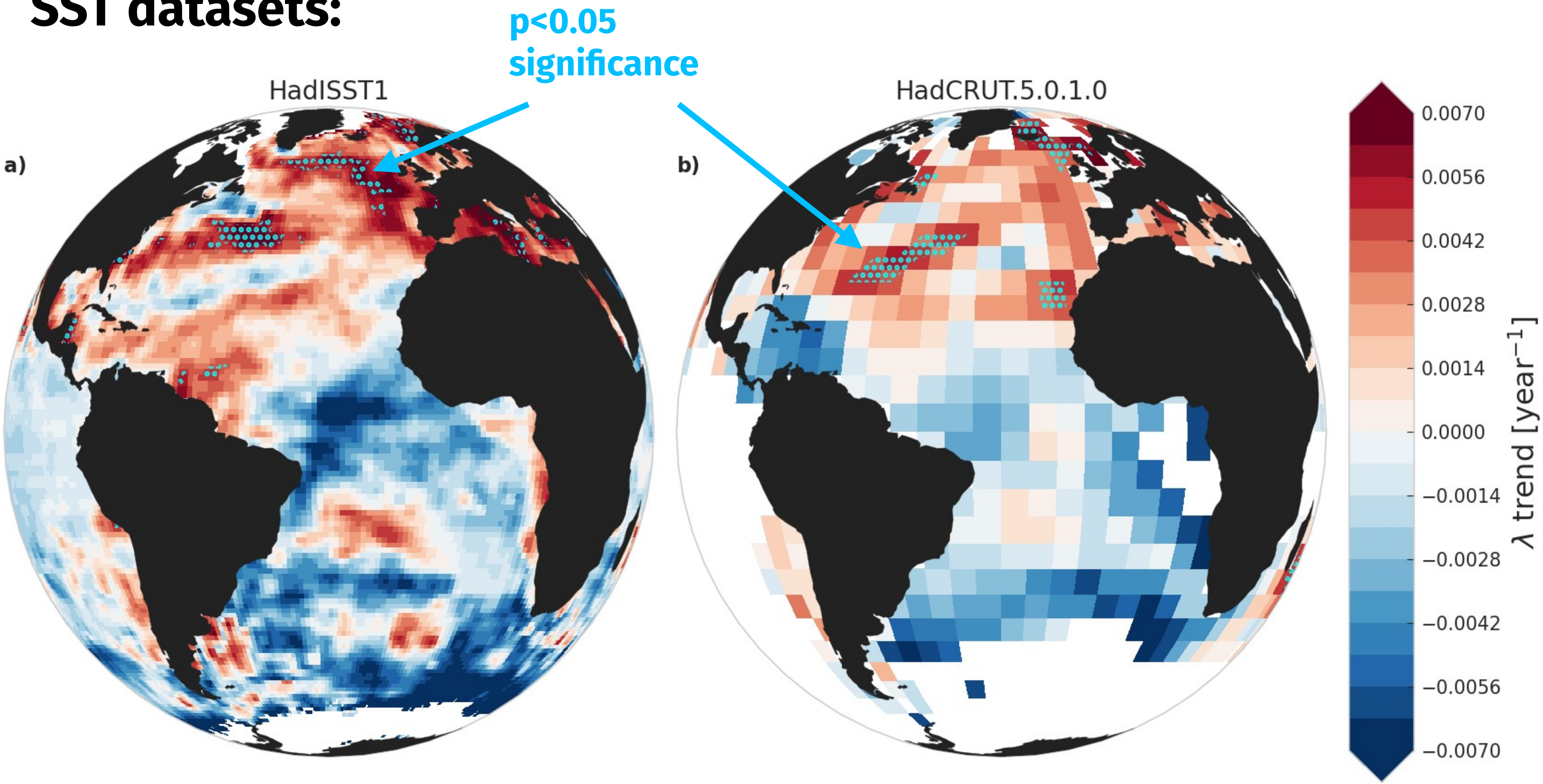
Overall, we use surrogates to calculate significant regions for three different datasets

HadISST1 Fourier surrogates

HadCRUT5 AR(2) surrogates

EN4.2.2 AR(2) surrogates (using the last 40 years)
+ modify to emulate analysis method

SST datasets:



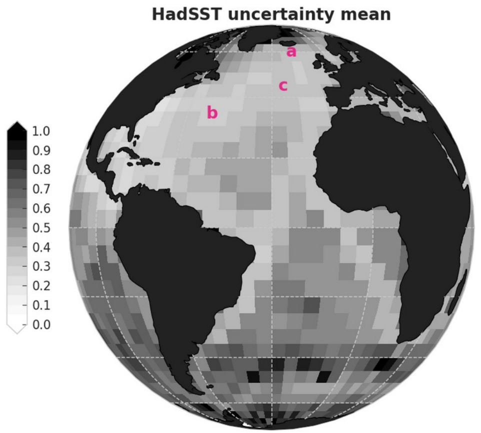
2 We use the uncertainty provided with the HadCRUT5 and HadSST4 datasets to estimate uncertainties in the AMOC sub-polar gyre (SPG) index.

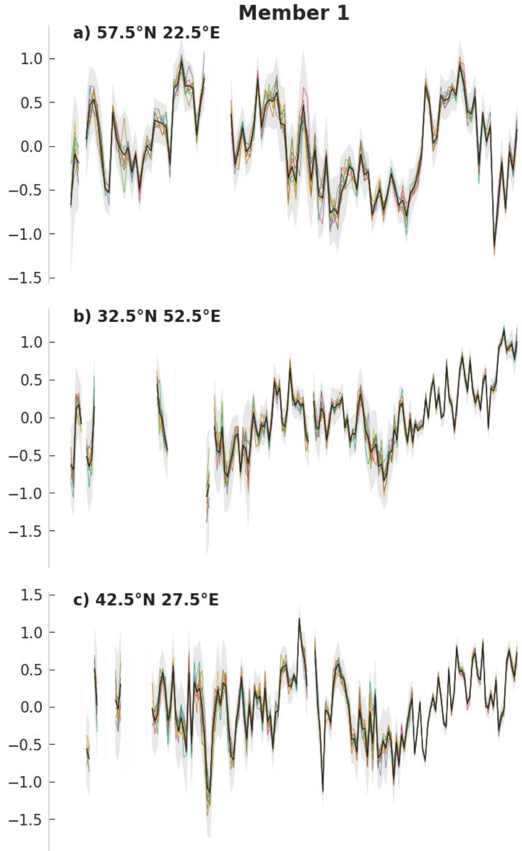
→ sample the uncertainty range at each gridpoint

→ use samples to calculate EWS in AMOC index

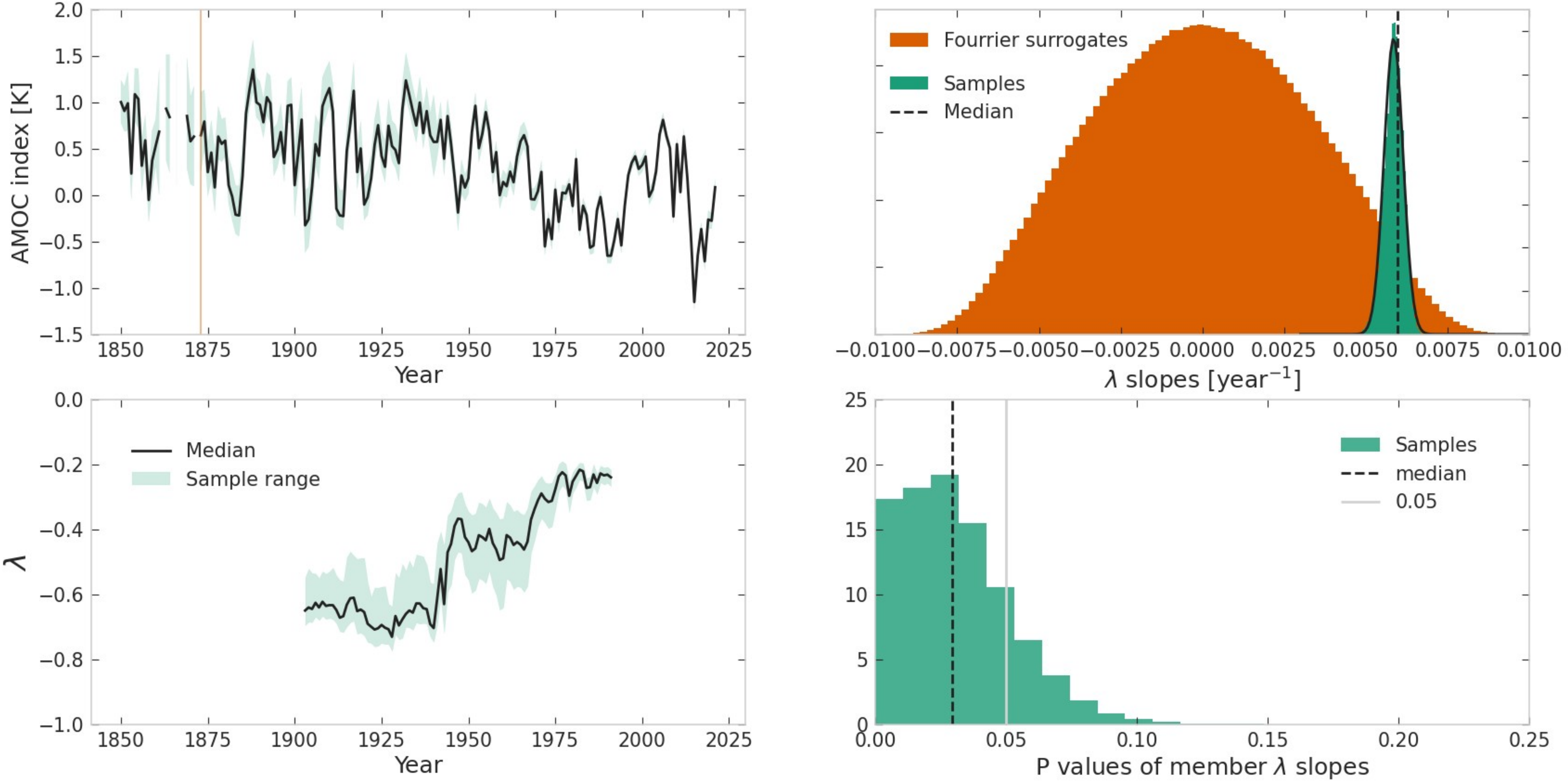
→ generate surrogates for each sampled index, and use these to calculate the significance of the trend

→ get distribution of p values for all the samples

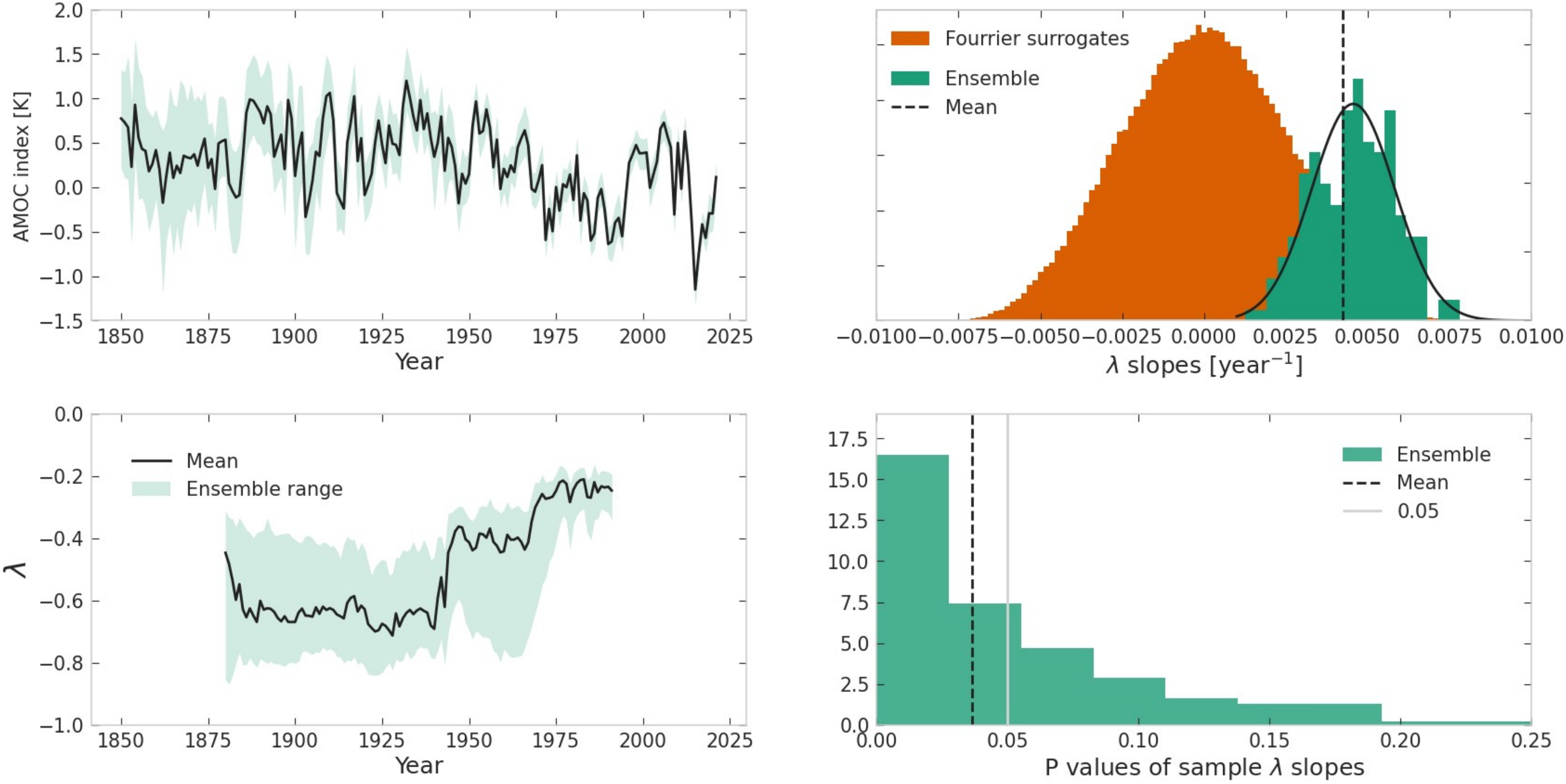




HadSST4 - index



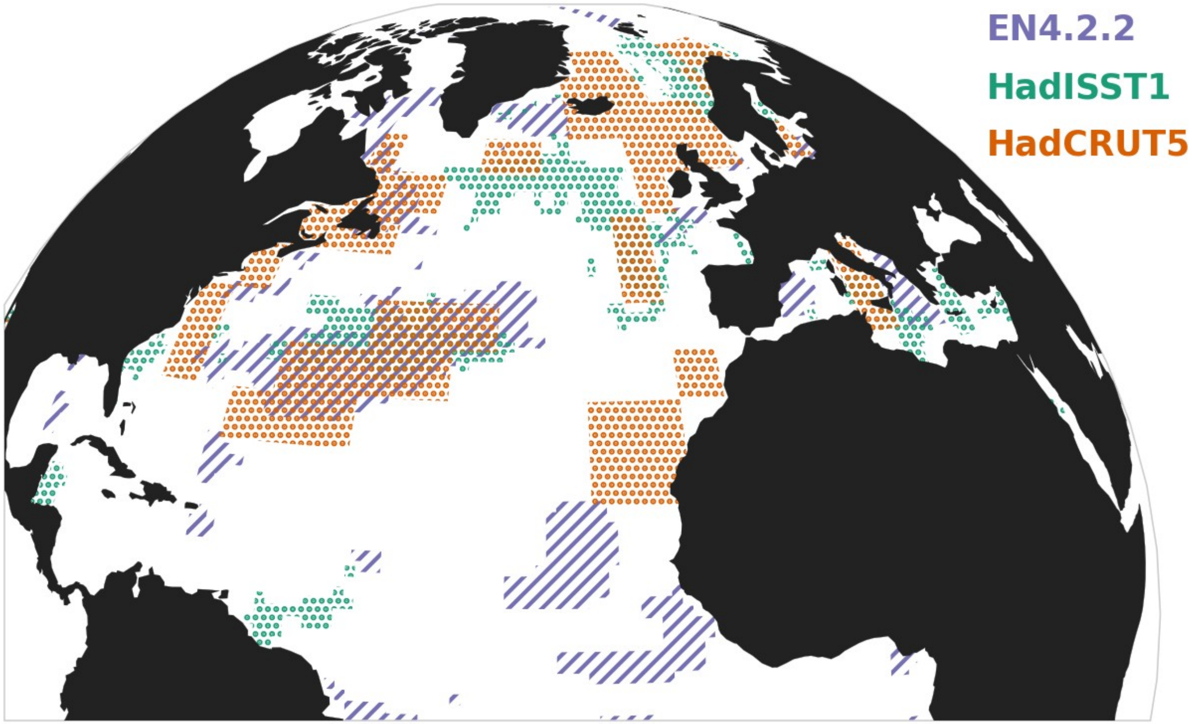
HadCRUT5 - index



Finally, we can look at where the regions of significant restoring rate increase are located in the North Atlantic for all three spatial datasets.

- southern edge of the North Atlantic Current
- In the Greenland, Iceland and Norwegian seas

→ these are regions typically associated with AMOC transport



Conclusions:

- **It’s important to take the dataset properties into account when calculating higher order statistics!**
- North Atlantic significant EWS still present

Preprint: arxiv.org/abs/2303.06448