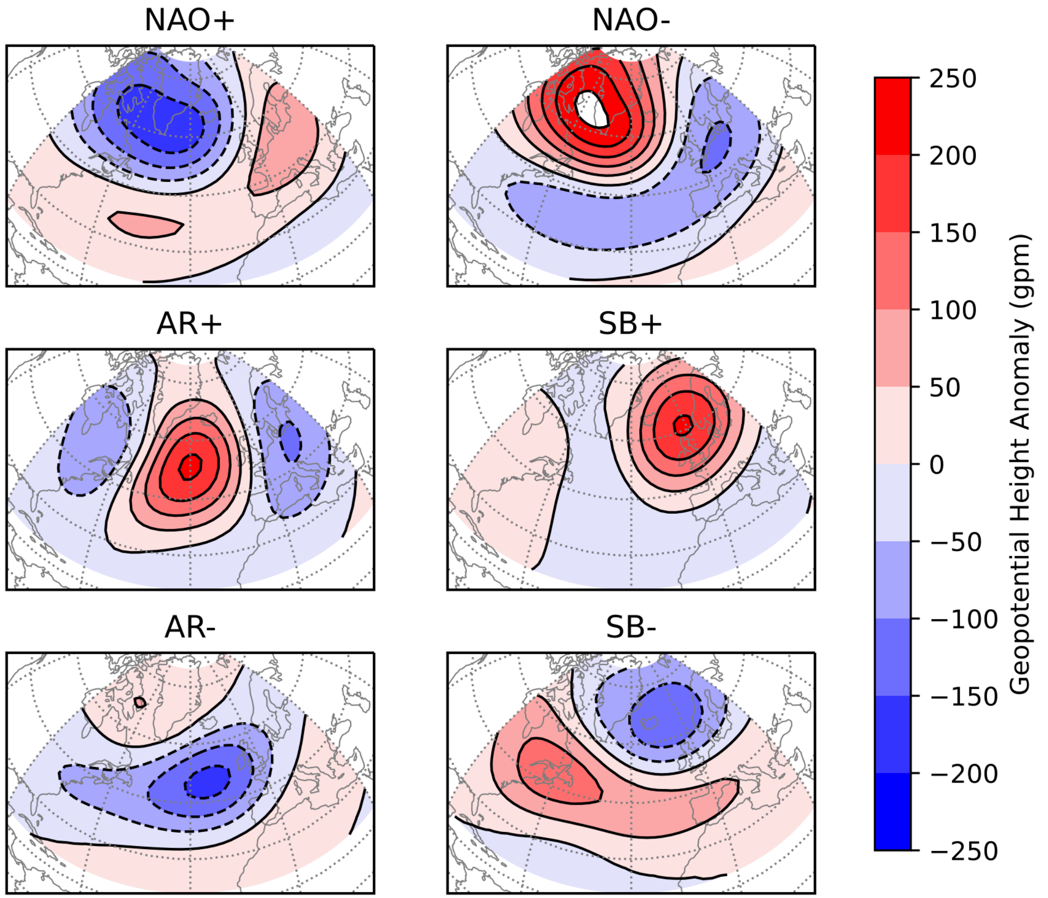
Non-Stationarity of Wintertime Atmospheric Circulation Regimes in the Euro-Atlantic Sector

Swinda Falkena[1], Jana de Wiljes[2],

Antje Weisheimer[3], Ted Shepherd[1]

1. University of Reading, 2. University of Potsdam, 3. ECMWF

Contact: s.k.j.falkena@pgr.reading.ac.uk @SwindaFalkena



EPSRC Centre for Doctoral Training

Imperial College London

University of Reading

Aim: Detect a robust non-stationary signal from ensemble data using circulation regimes

Difficulty: Models are imperfect, the regimes are domain dependent and exhibit a wide spread in regime frequencies

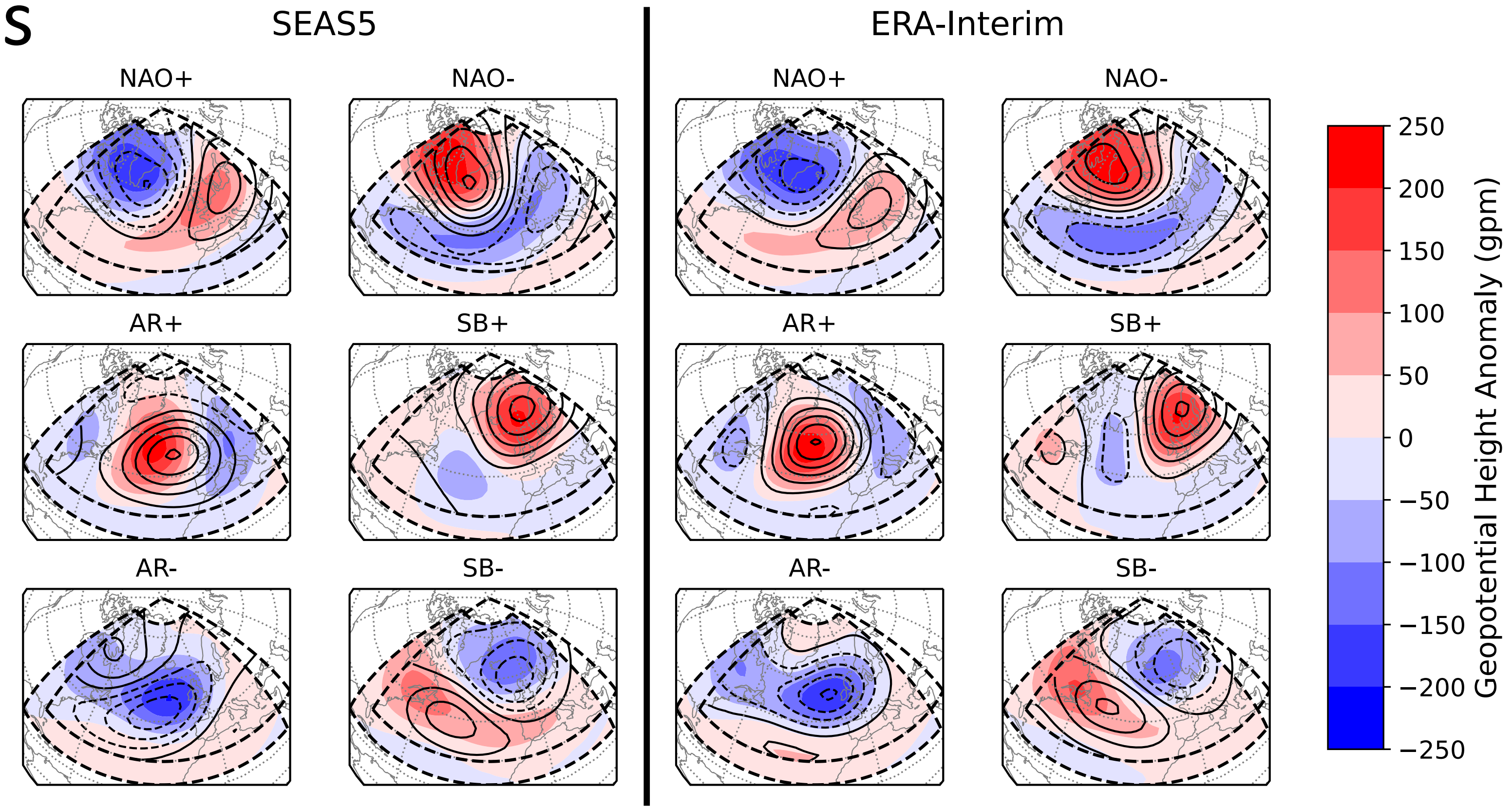


Fig: Regimes for SEAS5 (left) and ERA-Interim (right) for two different domains (dashed boxes) indicated by the colours and contours (same interval) respectively.

Solution: Use a regularized clustering method that enforces a level of similarity between ensemble members to identify the circulation regimes

Inter-annual variability

Linear regression shows **predictability for NAO+ and SB-** with a coefficient of 1, no signal for NAO-

Regression of an NAO-index yields a coefficient of 2

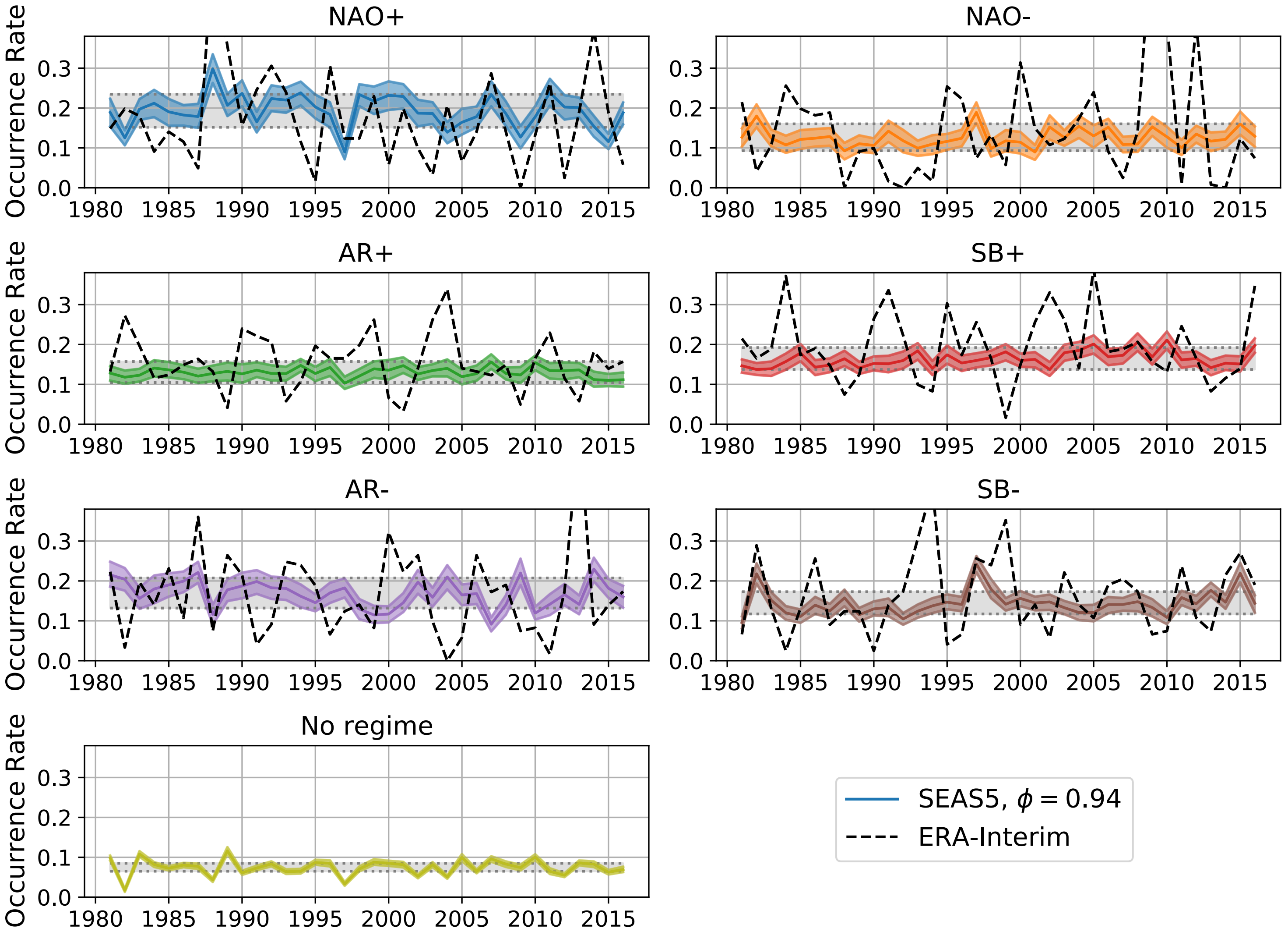


Fig: Inter-annual variability of regime occurrence rates for SEAS5 (colour) and ERA-Interim with the grey bars showing the noise level for SEAS5.

Similar signal strength for observations and model, possibly poor representation of NAO- links to **signal-to-noise paradox** for NAO-index

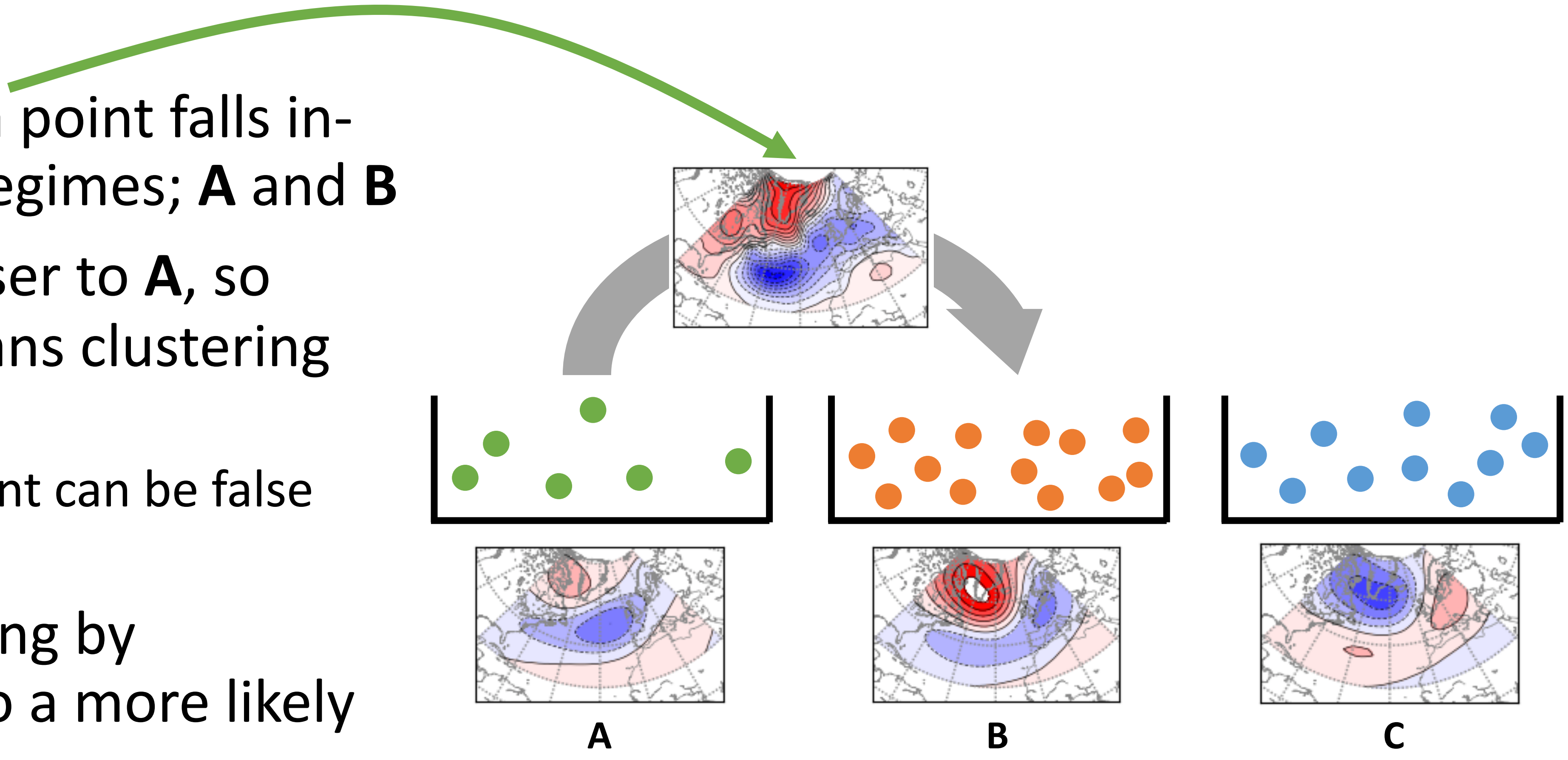
Display Contents

Data

- ECMWF SEAS5 hindcast ensemble
  - 51 members
  - November 1st start date
  - 1981-2016
- DJFM daily 500 hPa geopotential height (Z500)
- Domain A: 20-80°N, 90°W-30°E and Domain B: 30-90°N, 80°W-40°E
- Anomalies with respect to a constant climatology (DJFM average)
- Similar data of ERA-Interim reanalysis for comparison

Regularised k-means Clustering: The Idea

- At time *t* a data point falls in-between two regimes; **A** and **B**
- It is slightly closer to **A**, so standard k-means clustering assigns it there
  - This assignment can be false due to noise
- Detect overfitting by reassigning it to a more likely regime, i.e. **B**



Regularised k-means Clustering: The Maths

Clustering aims to split a data set into k clusters such that the within-cluster variance is minimised, but the between-cluster variance maximised. Let

- Ensemble data $x_{t,n} \in \mathbb{R}^{T\times N\times D}$ with T length of time series, N number of ensemble members and D dimension of the data (lat×lon) 1
- Cluster centres $\Theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^{k\times D}$ with k the number of clusters
- Affiliation vector $\Gamma = \big(\gamma_1(t,n), \dots, \gamma_k(t,n)\big) \in \mathbb{R}^{k\times T\times N}$ giving the assignment of data to the clusters

Minimize the averaged clustering functional

$$\mathcal{L}(\Theta,\Gamma) = \sum_t \sum_n \sum_k \gamma_k(t,n) g(x_{t,n}, \theta_k)$$

subject to

2 $$\sum_k \gamma_k(t,n) = 1$$

and the constraint

$$\textstyle\sum_k \sum_{n_1,n_2} |\gamma_k(t,n_1) - \gamma_k(t,n_2)| \le \phi \cdot C_{eq}$$

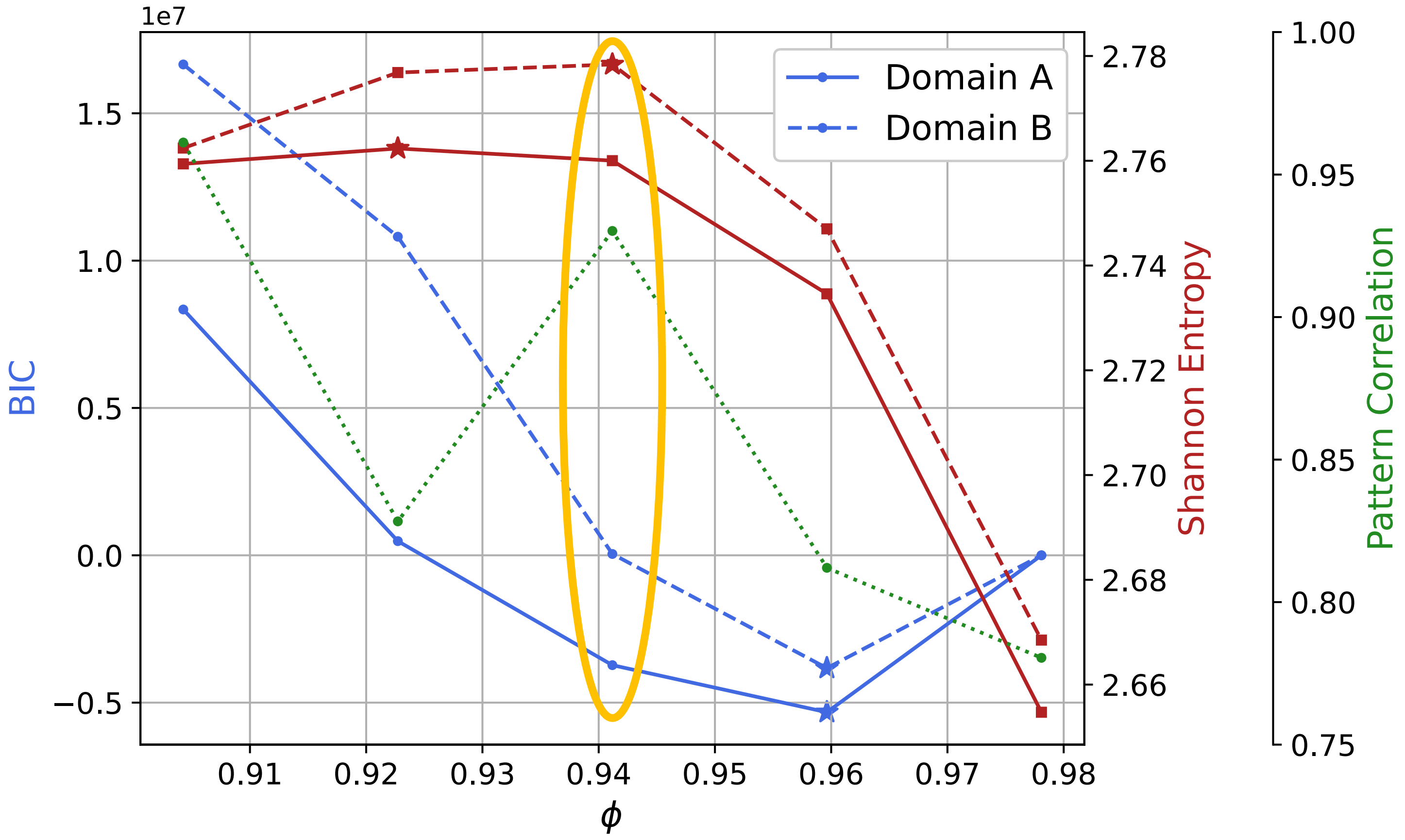
summing over all combinations of two ensemble members n_1, n_2 for every time t.

Identify a “no-regime” as data which cannot straightforwardly be assigned by the algorithm: 4

$$\gamma_k(t,n) \notin \{0,1\}$$

The constraint enforces the ensemble members to behave similarly at every timestep without making any assumptions on the form of the non stationarity 3

Regularised k-means Clustering: Selecting ϕ



Selection criteria for identifying a suitable constraint value:

- Bayesian Information Criterion (BIC)
  - Balance complexity and accuracy
- Shannon entropy
  - Information maximisation
- Domain robustness
  - Pattern correlation between domain A and B

Select the constraint value which allows to discriminate best between the regimes (high entropy) without losing reliability (low BIC): $\phi = 0.94$

Effect of Regularisation: Occurrence Rates

The overall occurrence rates of the regimes are more distinct, indicating the regularisation helps to better discriminate between regimes

1

The uniformity of the ERA-Interim occurrence rates is potentially due to a lack of discrimination between the regimes

2

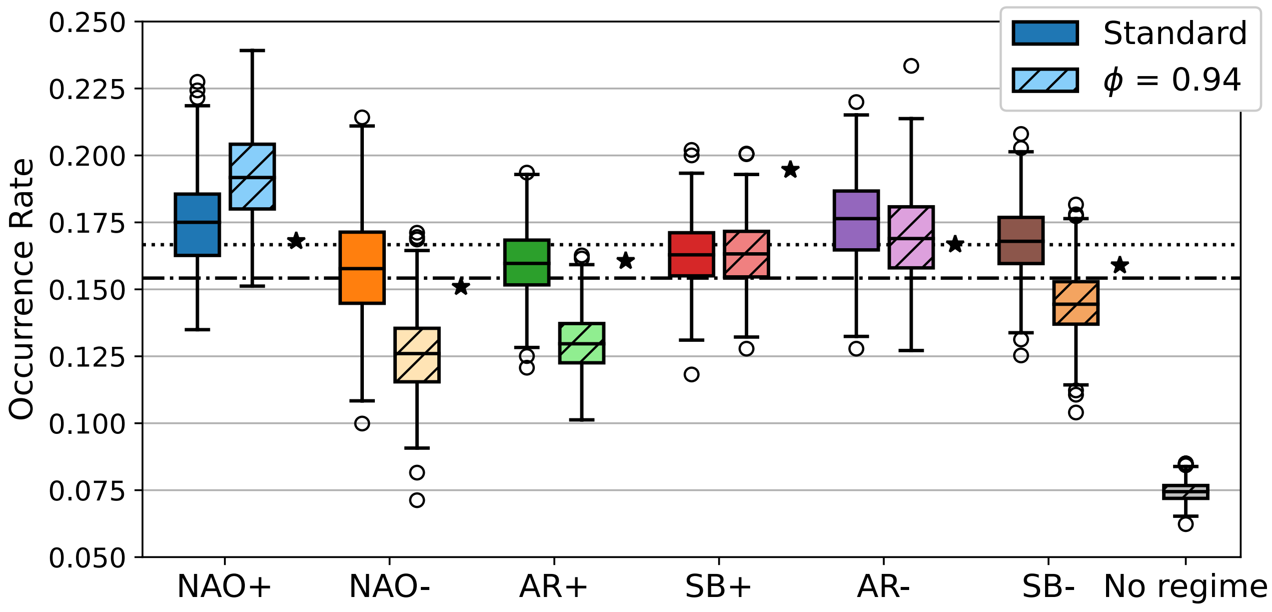


Fig: Occurrence rates for SEAS5 without (standard) and with constraint, where the dotted and dash-dotted lines, respectively, give represent an equal representation of the data over the regimes. The stars indicate ERA-Interim values.

Effect of regularisation: Data reassignment

Tab: Contingency table indicating the reassignment of data to a different regimes by the regularised algorithm.

The constrained regimes are no longer domain dependent

		$\phi = 0.94$							
		NAO+	NAO-	AR+	SB+	AR-	SB-	No-regime	Total
Unconstrained	NAO+	27994	0	1443	5367	54	1297	2934	39089
	NAO-	0	23301	87	37	9432	310	2069	36227
	AR+	14	3409	25867	2349	480	426	2981	35526
	SB+	412	1132	208	28005	3881	50	2670	36358
	AR-	12512	0	640	71	22190	795	2757	38965
	SB-	1946	104	722	685	1474	29265	3254	37450
	Total	42878	27946	28967	36514	37511	32143	16656	222615

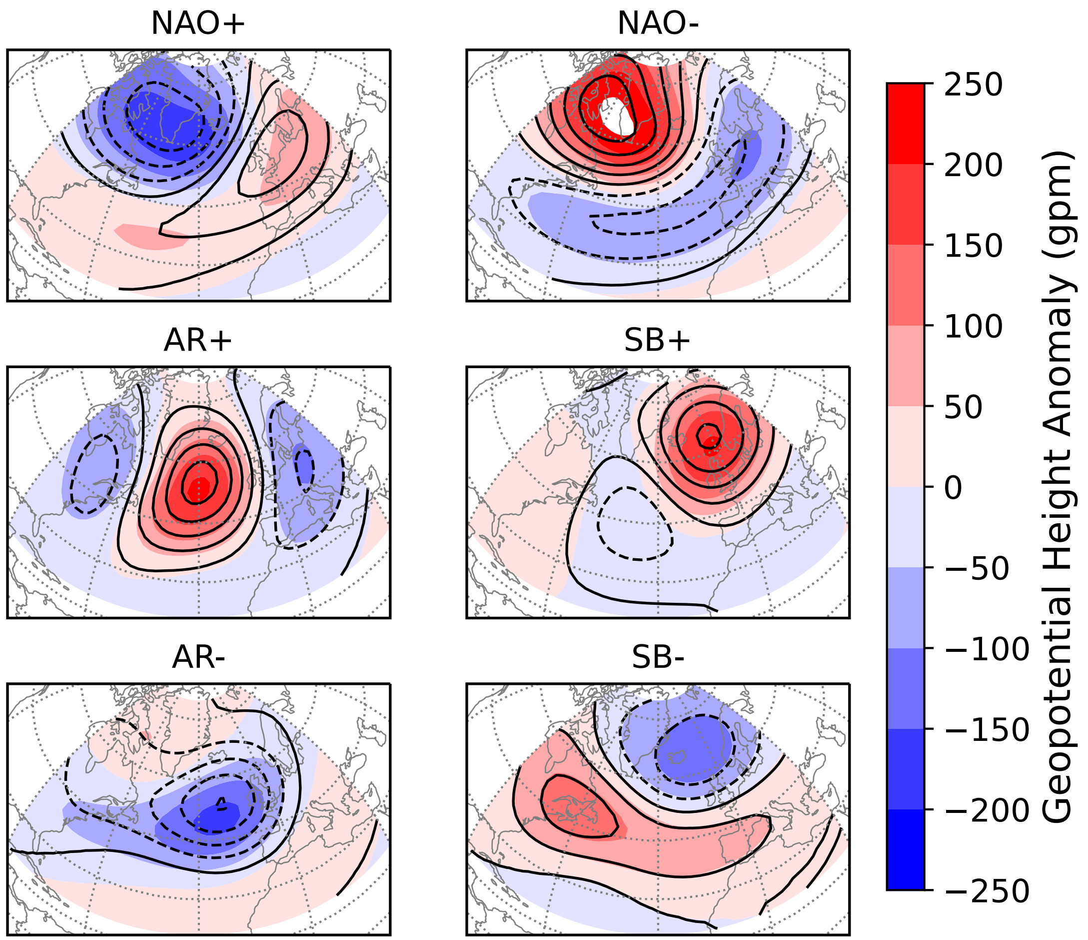
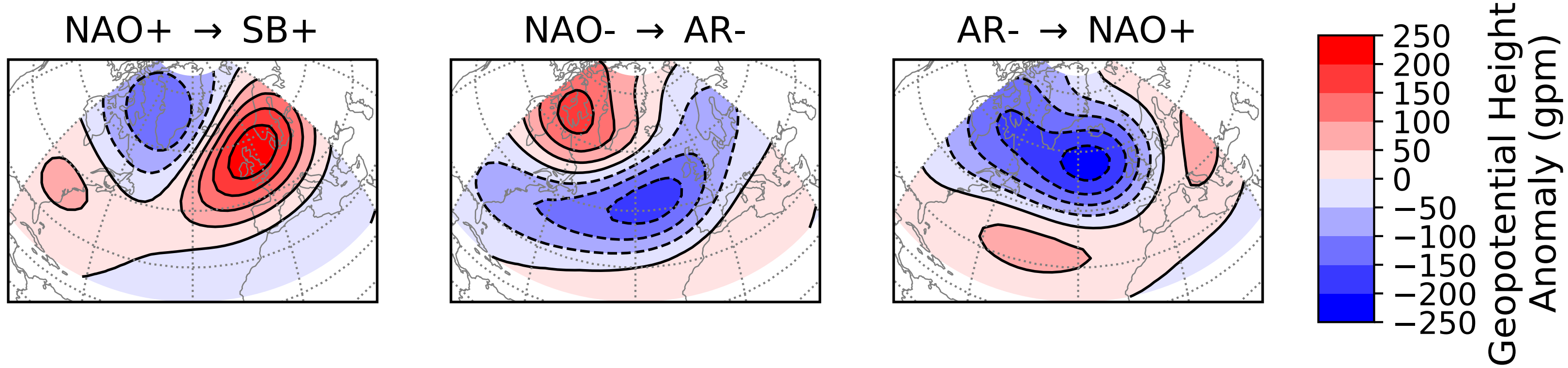
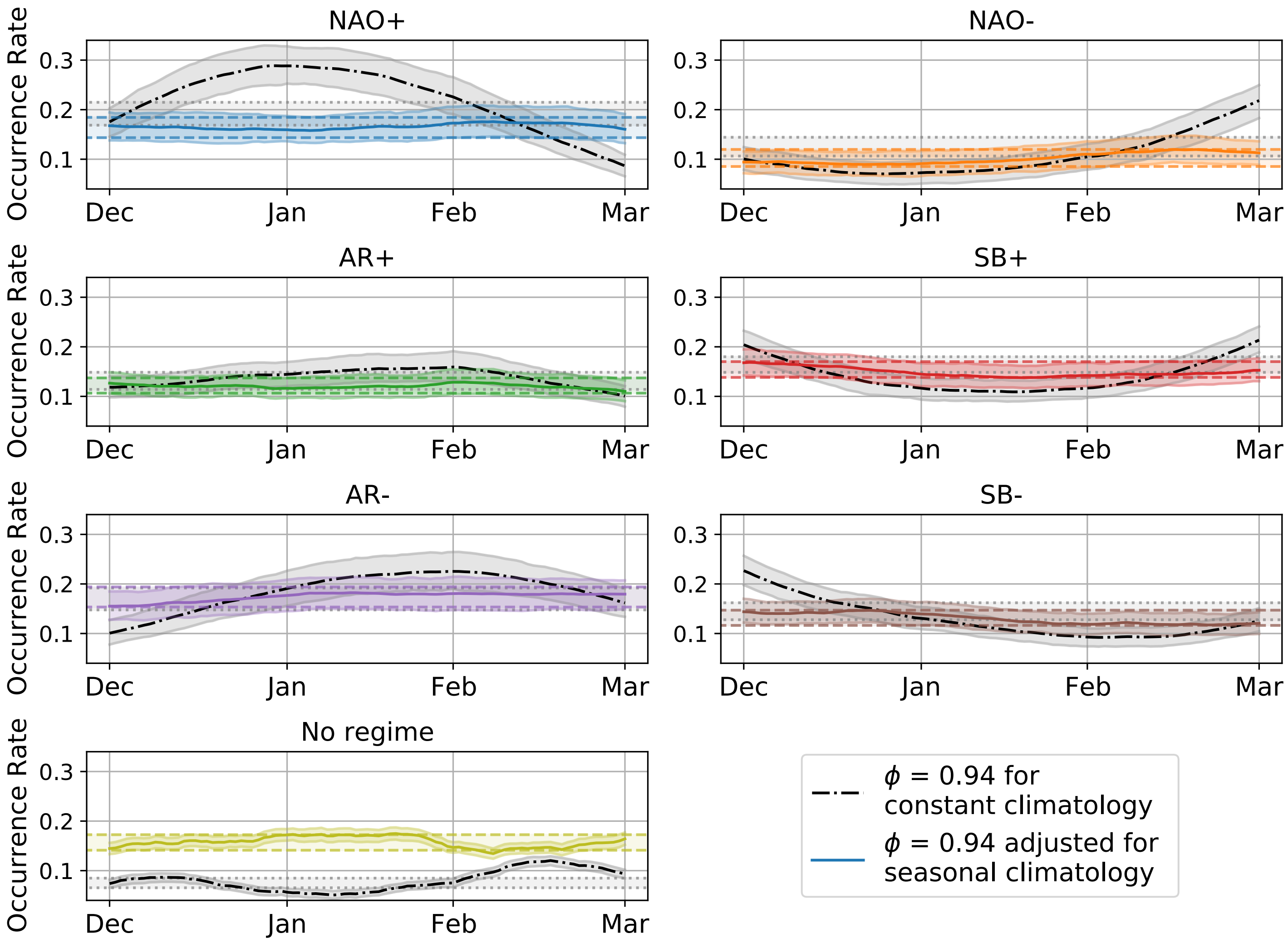


Fig ↑: Regimes for SEAS5 with (colours) and without (contours, same interval) constraint.



← Fig: Composites of data reassigned to a different regime by constraint for the cases indicated in red in the table.

Sub-seasonal Variability



1 A seasonal cycle in the occurrence rates is found

2 Adjusting for the seasonal cycle of the mean climatology (daily averages fitted with 4th-order polynomial) nearly all variability disappears

Inter-annual Variability

Strong inter-annual signals are found for NAO+, AR- and SB-, whereas the NAO- signal is weaker

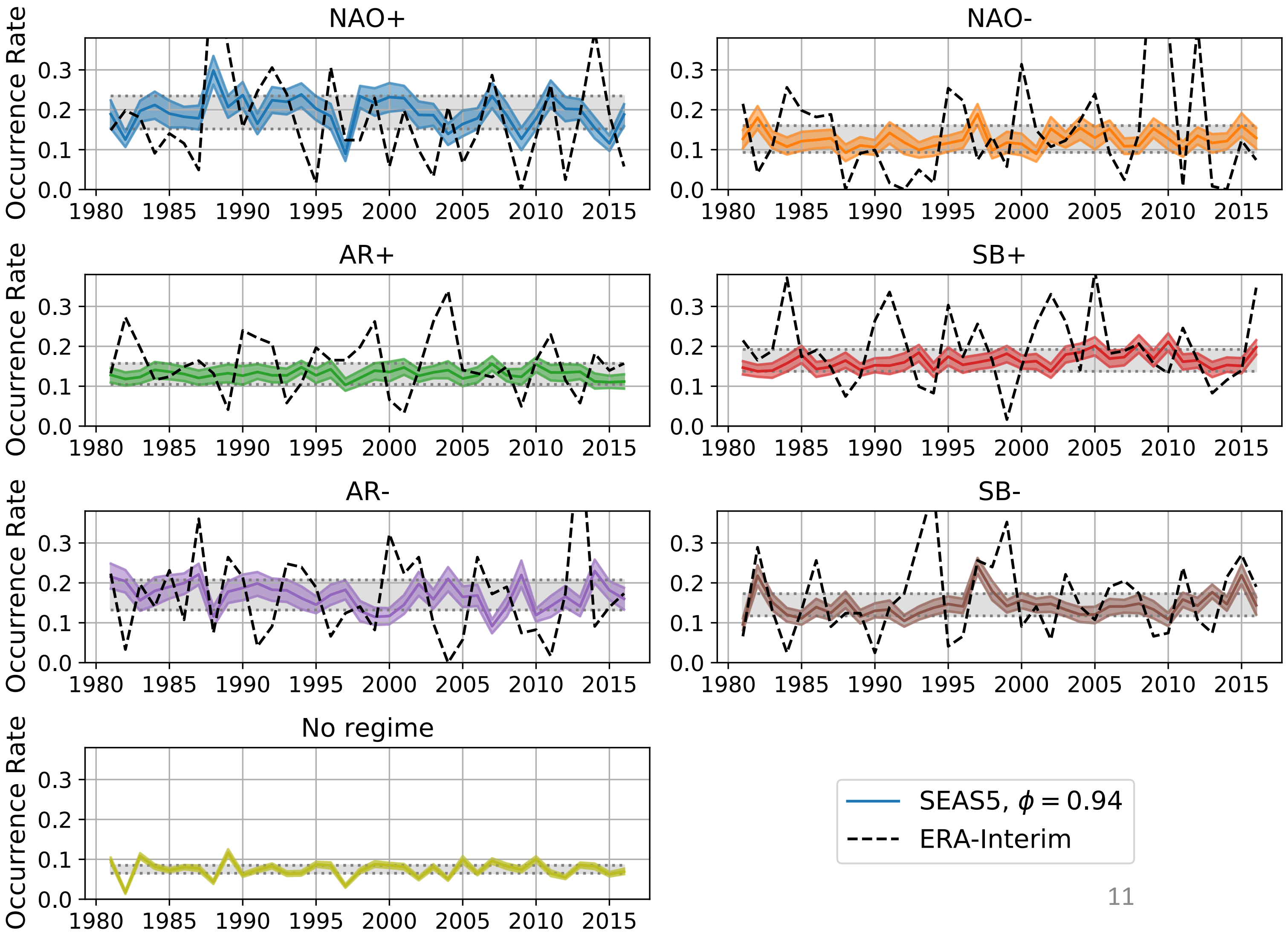
1

The majority of the signal coincides with El Niño and La Niña years, with NAO+ being less and SB- (and NAO-) more frequent

2

Note: most data assigned to both NAO+ and SB- would be assigned to NAO+ when considering only 4 regimes, i.e. 6 regimes allows to pick up a more detailed ENSO response

3



Inter-annual Variability: Regression Analysis

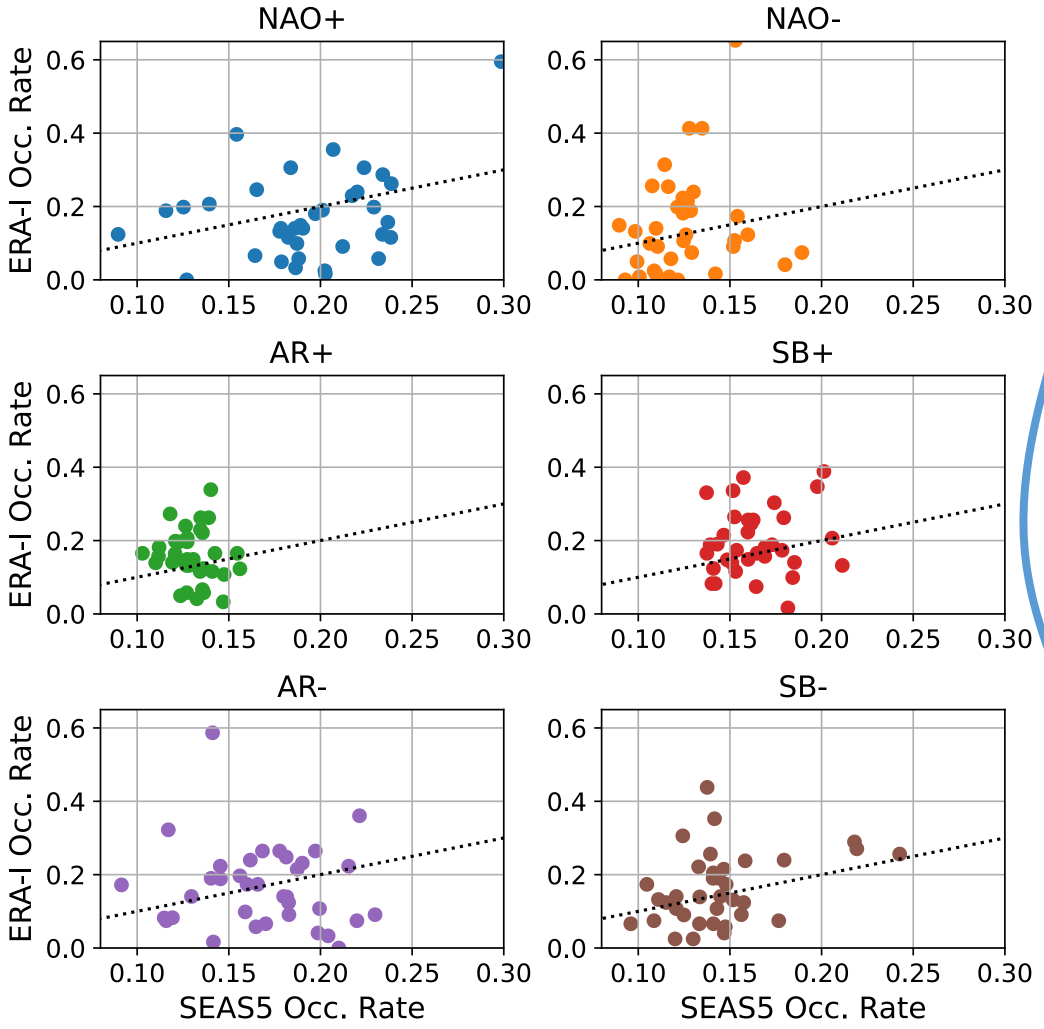


Fig: Scatter plots of annual regime occurrence with the dotted line showing a one-to-one relation.

Regime	NAO+	NAO-	AR+	SB+	AR-	SB-	MLR	NAO-
Slope	0.994	0.980	-0.685	0.530	-0.318	1.139	NAO+ SB-	-1.3481 -1.762
R^2	0.114	0.026	0.013	0.014	0.009	0.134		0.199
p	0.044	0.351	0.500	0.491	0.581	0.028		0.025
Bayes Factor	8.809	1.596	1.276	1.290	1.178	13.199		54.099

Tab: Results of a linear regression analysis for regime occurrence rates.

Predictable signals are found for NAO+ and SB- with a slope around 1

1

The Bayes factor indicates whether a regression hypothesis is more likely (>1) than a constant occurrence rate, the larger the better

Using multiple linear regression a strong signal for NAO- can be obtained from NAO+ and SB-

2

Inter-annual Variability: Signal-to-Noise

No signal-to-noise paradox is found for NAO+ and SB-occurrence rates 1

Linear regression for an NAO-index indicates that here the model is better at predicting observations than itself with a coefficient of 2 2

A poor model representation of NAO- could be linked to the signal-to-noise paradox for the NAO-index 3

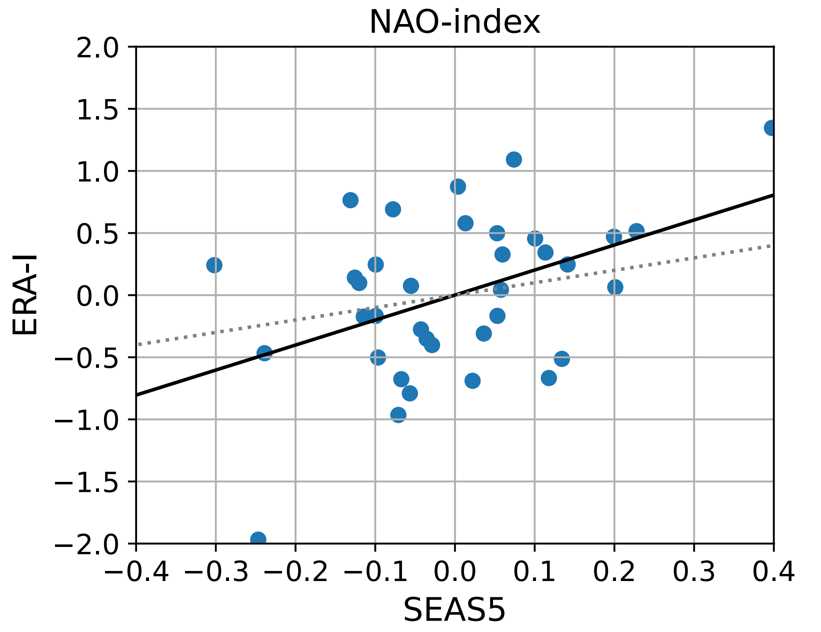


Fig: Regression analysis for an NAO-index with a slope of 2, with the dotted line giving the one-to-one reference line.

Regression analysis can be used to identify the signal strength:

- Assume a true signal $c(t)$
- Observational time series $y(t) = a\, c(t) + e_y(t)$, with a the signal strength and $e_y(t)$ noise
- Similarly for an ensemble member $x_i(t) = b\, c(t) + e_{x_i}(t)$ and the ensemble mean $\bar{x}(t) = b\, c(t) + e_{\bar{x}}(t)$

- Regression of $y(t)$ onto $\bar{x}(t)$ yields a coefficient of $^a/_b$
- $^a/_b > 1$ indicates the model is better at predicting observations than its own ensemble members
- Ratio of Predictable Components = $\frac{a}{b}\frac{\sigma_{x_i}}{\sigma_y}$, with $\sigma_{x_i,y}$ the standard deviations of the residuals

References

Falkena, S.K.J., de Wiljes, J., Weisheimer, A., Shepherd, T.G., Detecting robust non-stationary signals from the noise in ensemble data using circulation regimes, *in preparation*.

Falkena, S.K.J., de Wiljes, J., Weisheimer, A., Shepherd, T.G., Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector, *Quarterly Journal of the Royal Meteorological Society*, 2020.

Hannachi *et al.*, Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere, *Reviews of Geophysics*, 2017.

Fabiano *et al.*, Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance, *Climate Dynamics*, 2020.

Eade *et al*., Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, *Geophysical Research Letters*, 2014.