The predictability of sea surface temperatures in El Niño regions

Tommaso Proietti
Università di Roma "Tor Vergata"
Alessandro Giovannelli
Università dell'Aquila

High Dimensional Data: theory and applications
(A synergy between physics, statistics and economy)
Laboratori Nazionali del Gran Sasso
30 Novembre 2023

Introduction

- El Niño/Southern Oscillation (ENSO) phenomenon is one of the most important sources of interannual climate variability.
- It results from the interaction of an oceanic component and an atmospheric component.
- El Niño refers to the periodic warming of waters near the South-American equatorial Pacific coast around Christmas.
- The atmospheric component is evidenced by the difference in atmospheric surface pressure in the western (Darwin) and eastern (Tahiti) Pacific.
- When sea surface temperatures (SST) are high in the eastern pacific (positive anomaly), pressure is lower than normal, relative to the west (negative anomaly), and the Southern Oscillation Index (SOI) shows a negative anomaly.
- The pressure gradient produces anomalous winds blowing from west to east.

- Also, positive subsurface temperature anomalies are observed in the west. The thermocline, separating the upper from the deep ocean, is deeper than normal in the east: sea surface height and pressure in the upper ocean are higher than normal.
- Above is a description of an extreme phase. The opposite phase is called La Niña, during which the sign of the anomalies is reversed.
- An excellent description of the transition mechanism from one phase to the other is provided by Neelin (2011).

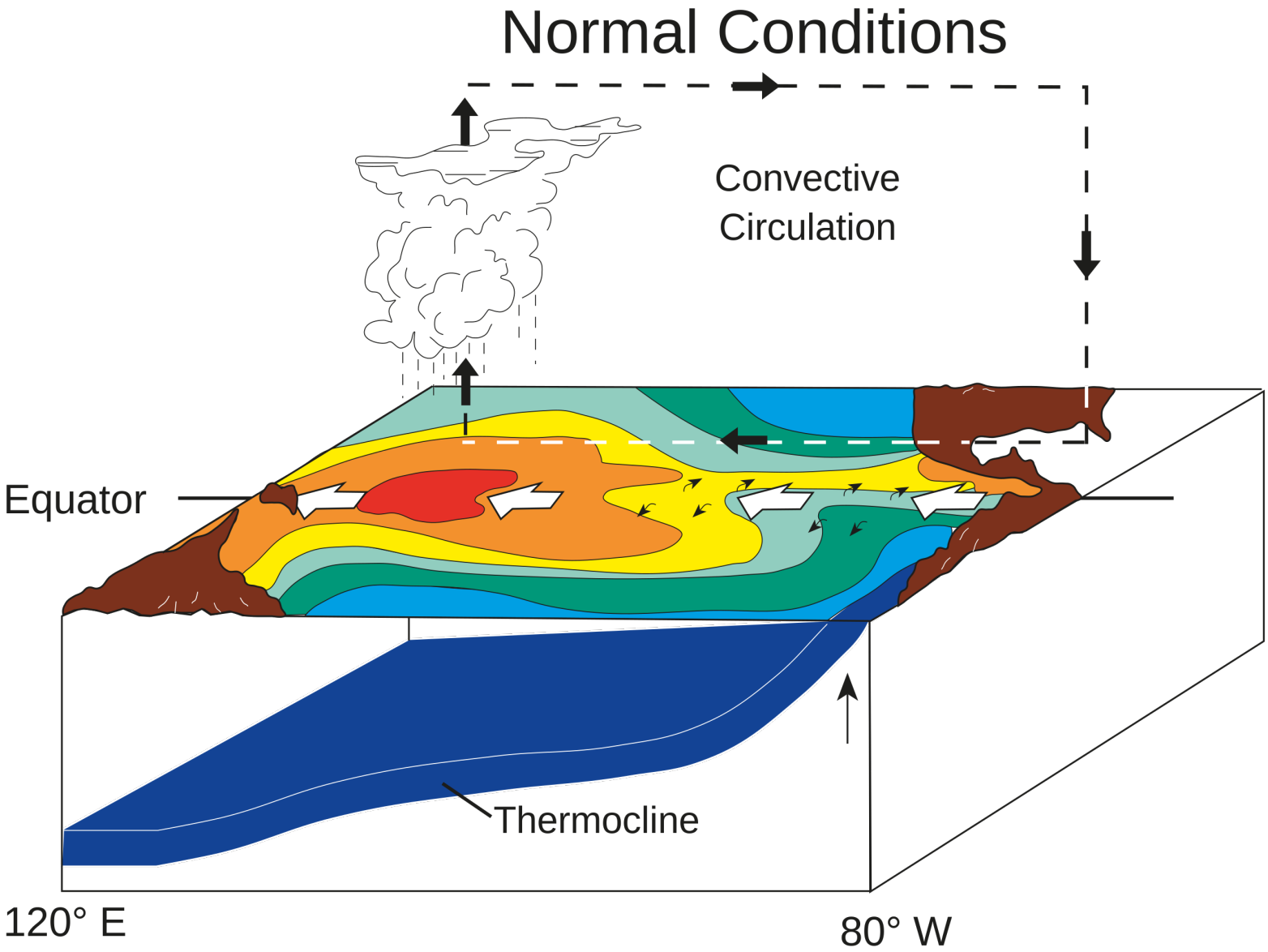


Figure: Illustrations from What is an El Nino?
Courtesy of the TAO Project at NOAA/Pacific Marine Environmental Laboratory.

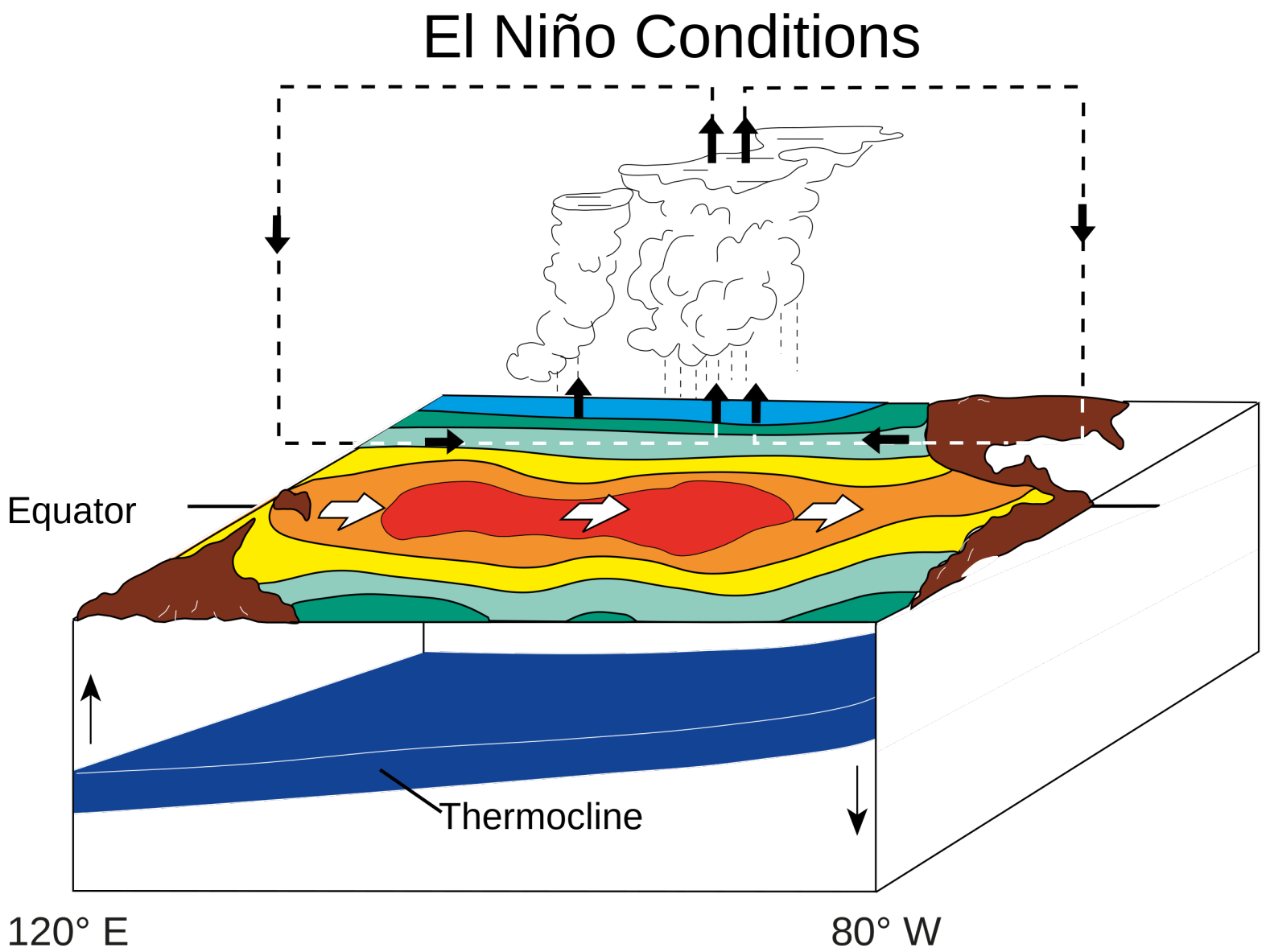


Figure: Illustrations from What is an El Nino?
Courtesy of the TAO Project at NOAA/Pacific Marine Environmental Laboratory.

- Although it originates in the tropical Pacific, ENSO's effects induce large weather and climate anomalies worldwide.
- ENSO is a predictable phenomenon. Interest lies in multistep prediction with horizon from 6 to 12 months.
- Two approaches are available: statistical (data driven, reduced form) models, and physical (coupled ocean-atmosphere) models.

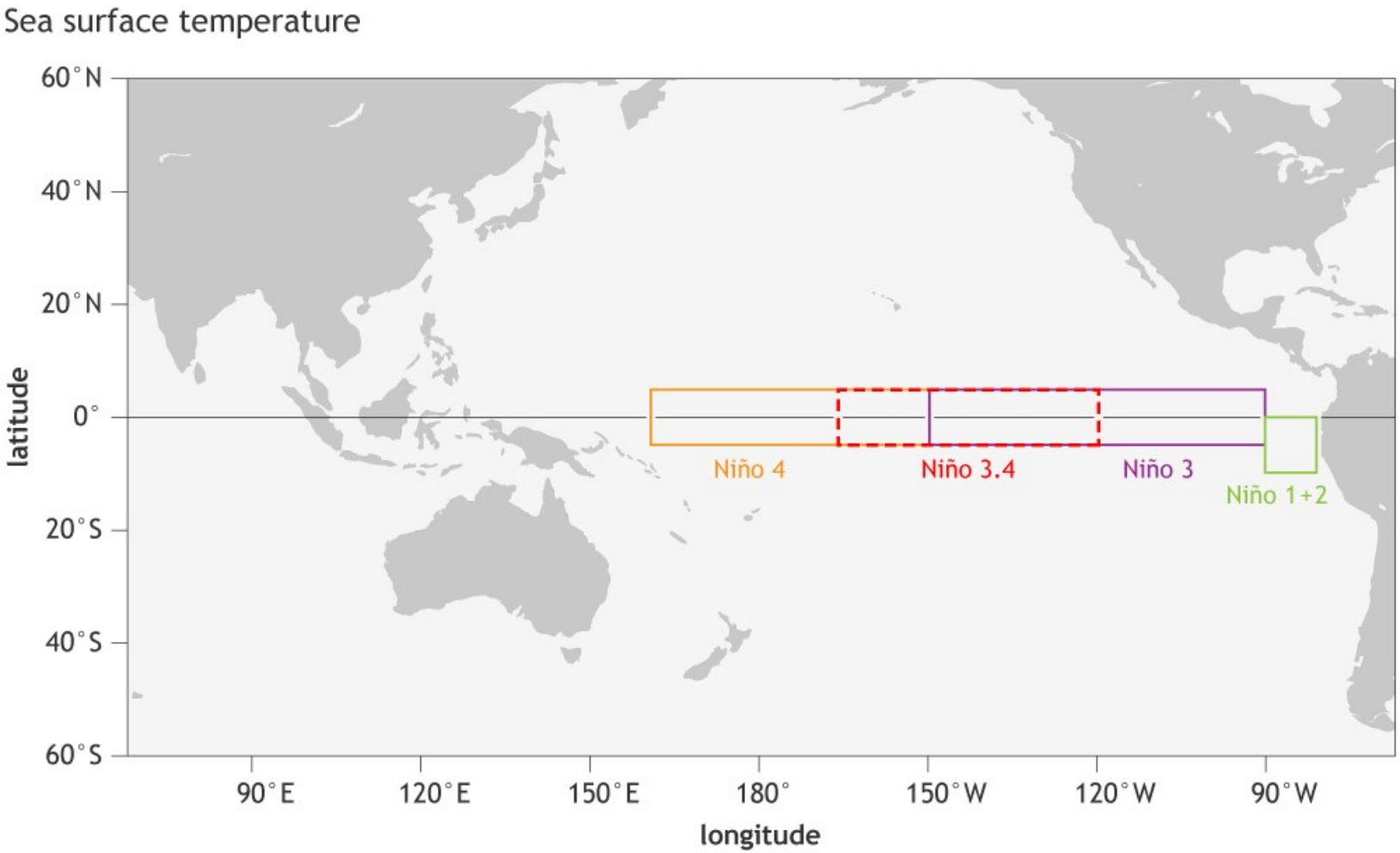
Description of the dataset

We consider a multivariate time series with $N = 15$ series (Source: NOAA https://www.cpc.ncep.noaa.gov/data/indices/) for the period Jan. 1980 - July 2023

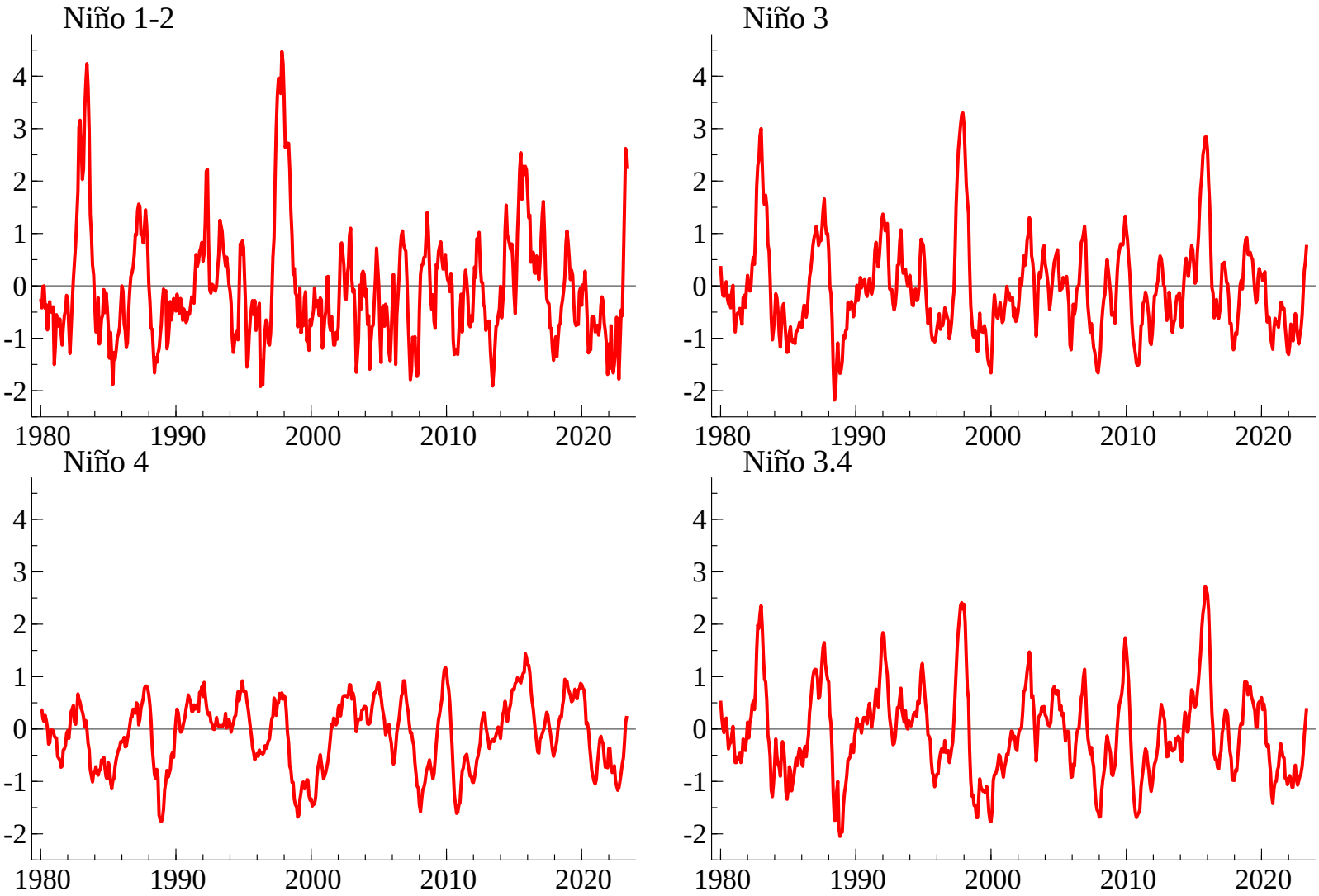
- Sea Surface Temperatures (Monthly ERSSTv5, 1991-2020 base period) : 4 time series

Niño 1+2	(0-10° South)	(90°West-80°West)
Niño 3	(5°North-5°South)	(150°West-90°West)
Niño 4	(5°North-5°South)	(160°East-150°West)
Niño 3.4	(5°North-5°South)	(170°West-120°West)

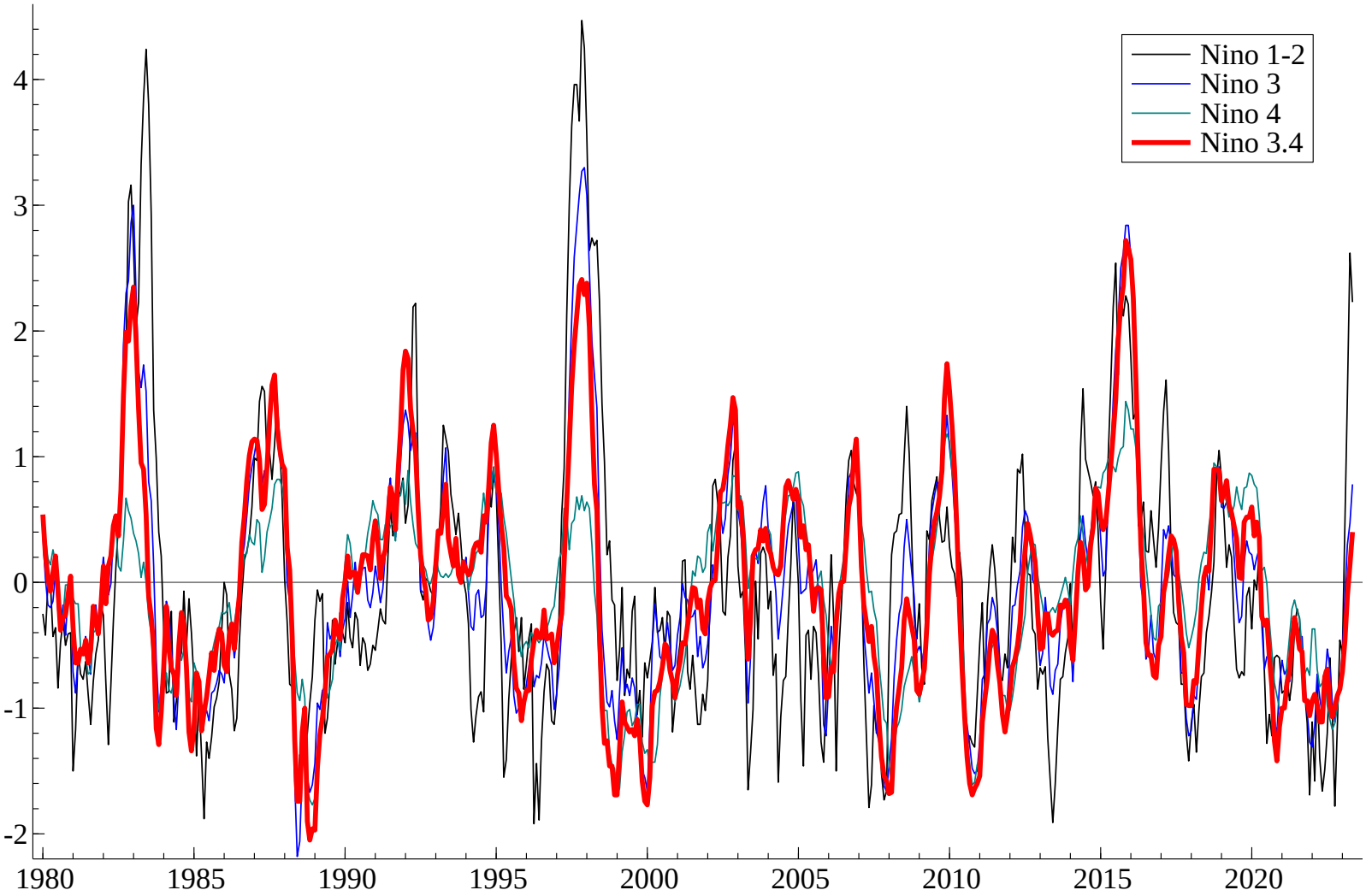
The phases are often characterized on the basis of the Niño 3.4 series. This will represent our target time series.


Sea surface temperature
60°N
40°N
20°N
0°
20°S
40°S
60°S
latitude
90°E
120°E
150°E
180°
150°W
120°W
90°W
longitude
Niño 4
Niño 3.4
Niño 3
Niño 1+2

SST series

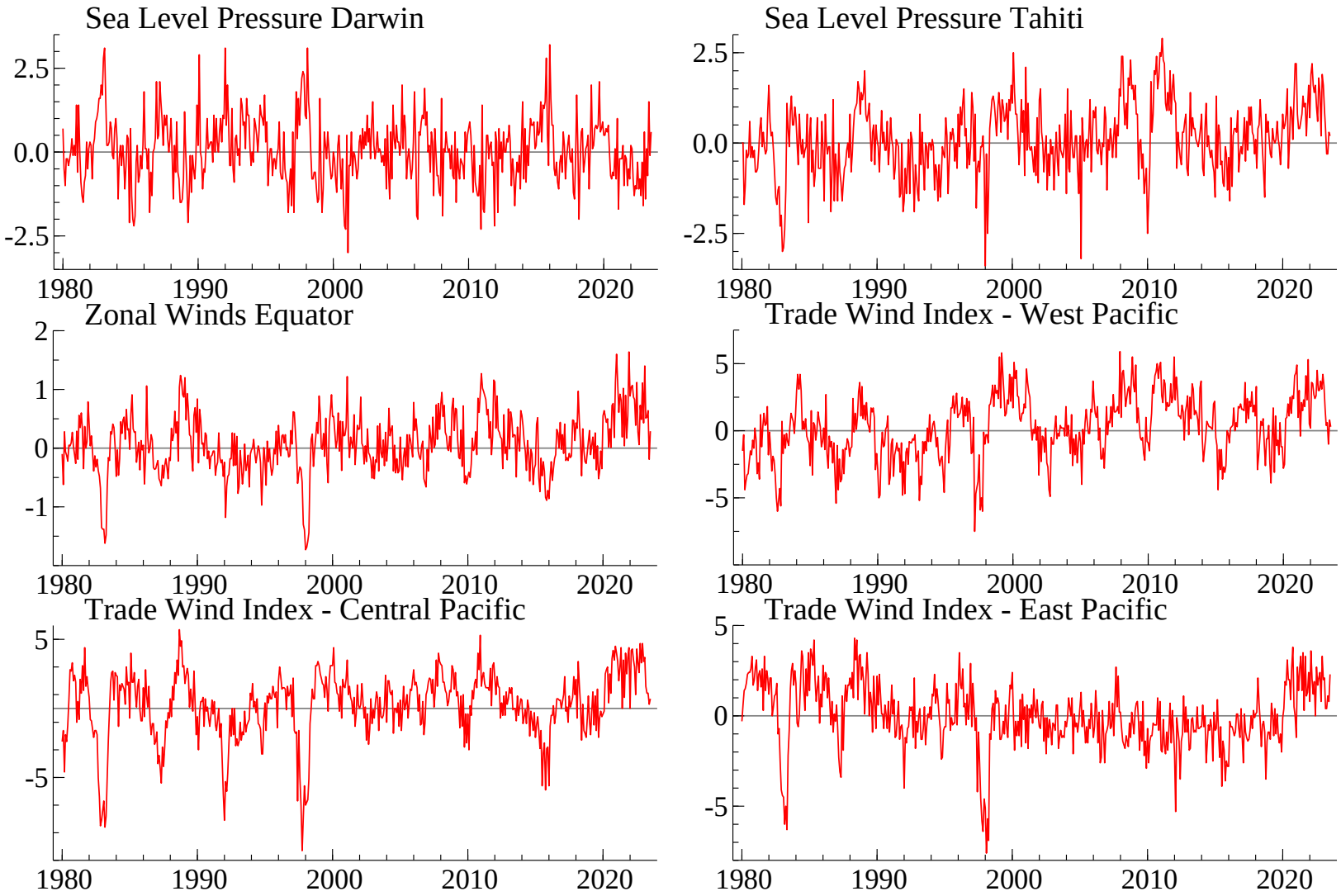


SST series (comparison)



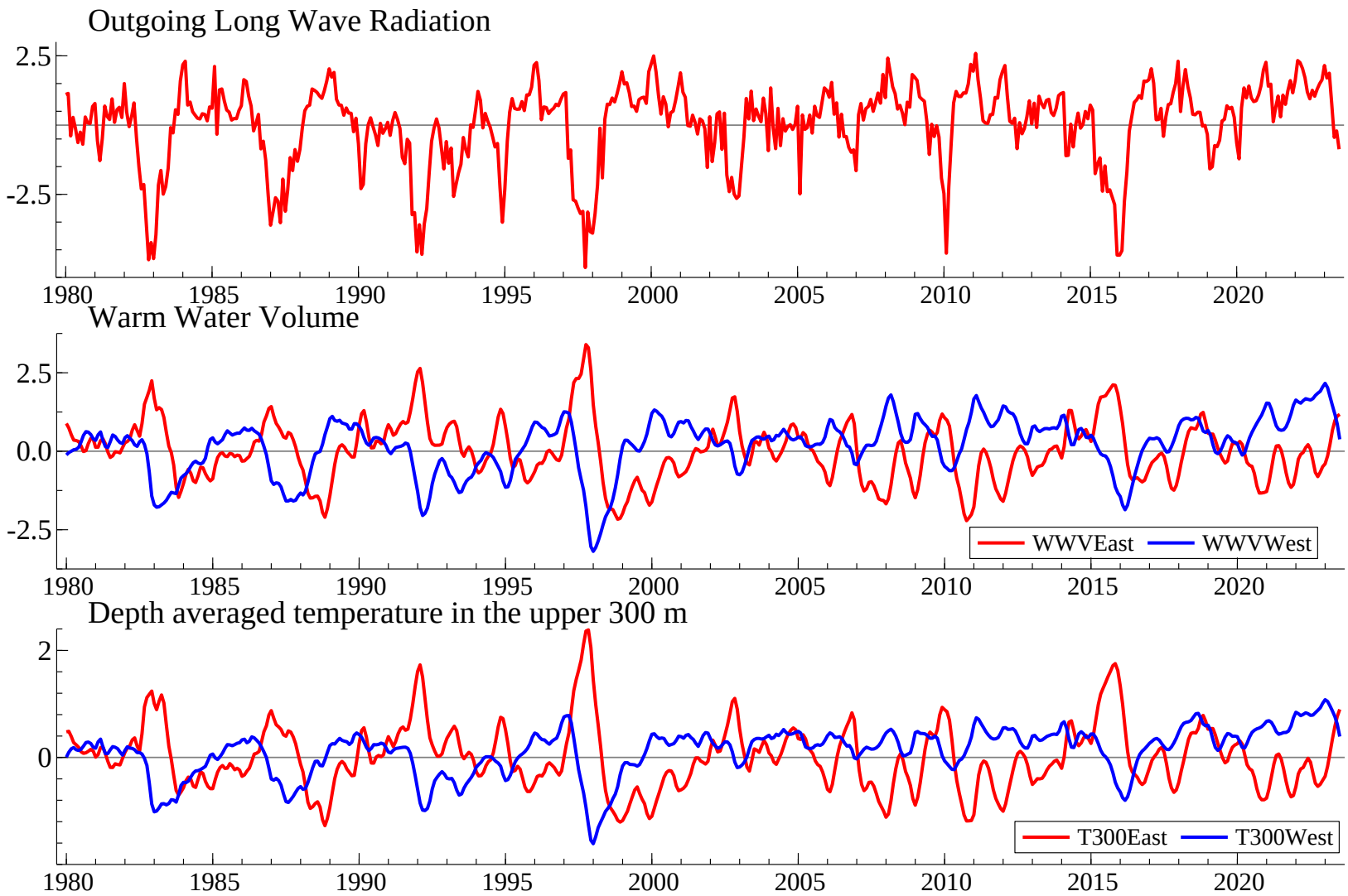
- Sea level pressure anomalies
  - SLP Tahiti (149.6 W, 17.5 S).
  - SLP Darwin (130.9 E, 12.4 S).
- Winds
  - ZWEq: 200 millibar (mb) Zonal Winds Equator (165°West-110°West)
  - WPTWI: 850 mb Trade Wind Index (5°North-5° South) (135°East-180°West) West Pacific
  - CPTWI: 850 mb Trade Wind Index (5°North-5° South) (175°West-140°West) Central Pacific
  - EPTWI: 850 mb Trade Wind Index (5°North-5° South) (135°West-120°West) East Pacific

Sea level pressure and wind series



- Outgoing Long Wave Radiation
  - OLR: Monthly Central Pacific OLR Index (1991-2020 base period, 170°E-140°W,5°S-5°N)
    Negative (Positive) OLR are indicative of enhanced (suppressed) convection and hence more (less) cloud coverage typical of El Niño (La Niña) episodes.
- Upper Ocean Heat Content (https://www.pmel.noaa.gov/elnino/upper-ocean-heat-content-and-enso)
  - WWVEast: Warm Water Volume East Pacific (5°North-5°South) (155°West-80°West)
  - WWVWest: Warm Water Volume West Pacific (5°North-5°South) (120°East-155°West)
  - T300East: depth averaged temperature in the upper 300 m (5°North-5°South) (155°West-80°West)
  - T300West: depth averaged temperature in the upper 300 m (5°North-5°South) (120°East-155°West)

OLR and Upper Ocean Heat Content Series



Methodology: Assumptions

- Let $\{Y_t, t = 1, \ldots\}$ be a real valued, N-dimensional stationary random process, with mean zero and cross-covariance function

$$\Gamma(k) = \mathrm{E}(Y_t Y'_{t-k}), j = 0, \pm 1, \ldots,$$

  with elements $\gamma_{ij}(k) = \mathrm{E}(Y_{it} Y_{j,t-k})$, so that $\Gamma(-k) = \Gamma(k)'$.
- We assume that $\sum_{k=-\infty}^{\infty} k\|\Gamma(k)\| < \infty$.
- We denote by

$$\Gamma_n = \begin{bmatrix} \Gamma(0) & \Gamma(1) & \cdots & \Gamma(n-1) \\ \Gamma(-1) & \Gamma(0) & \ddots & \Gamma(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ \Gamma(1-n) & \Gamma(2-n) & \cdots & \Gamma(0) \end{bmatrix}, \qquad (1)$$

  the $nN \times nN$ cross-covariance matrix of $\{Y_1, \ldots, Y_n\}$.

- The spectral density matrix,

$$F(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) e^{-\imath \omega j}, \quad \omega \in (-\pi, \pi], \tag{2}$$

  where $\imath^2 = -1$, is Hermitian and positive definite.
- We consider the estimation of Γ_n from a sample realization of length n, $\{y_t, t = 1, 2, \ldots, n\}$ and the related problems of estimating $F(\omega)$.
- We consider optimal linear prediction based on the finite past.
- The vector autoregressive (VAR) predictor with p lags entails the estimation of $225p$ AR coefficients.

A sketch of the approach: multivariate Durbin-Levinson

- The multivariate Durbin-Levinson (MDL) algorithm (Whittle 1963) processes the cross-covariance matrices

$$\{\Gamma(0), \Gamma(1), \ldots, \Gamma(m)\},$$

$m \leq n$, and outputs the matrices

$$\{\Phi_{m1}, \Phi_{m2}, \ldots, \Phi_{mm}; \Sigma_m, m = 1, \ldots, n\},$$

which are the coefficient matrices of the minimum mean square linear predictor (MMSLP) of Y_t given m past observations $\hat{Y}_{m,t} = \Phi_{m1} Y_{t-1} + \Phi_{m2} Y_{t-2} + \cdots + \Phi_{mm} Y_{t-m}$, and the one-step-ahead prediction error variance matrix.

Multivariate Durbin-Levinson Details

- From the output of the MDL algorithm we can derive the **partial cross-correlation**(PCC) matrices

$$\{\mathcal{C}(1), \mathcal{C}(2), \ldots, \mathcal{C}(m)\},$$

such that $\mathcal{C}(k)$ is the correlation matrix between the prediction errors $Y_t - \hat{Y}_{k,t}$ and $Y_{t-k} - \hat{Y}^*_{k,t-k}$, where $\hat{Y}^*_{k,t-k}$ is the MMSLP of Y_{t-k} based on $Y_{t-1}, \ldots, Y_{t-k+1}$.

- Consider the singular value decomposition

$$\mathcal{C}(k) = L_k \mathcal{P}_k R'_k,$$

where $L_k L'_k = L'_k L_k = I_N$, $R_k R'_k = R'_k R_k = I_N$, and $\mathcal{P}_k = \text{diag}\{\varrho_1(k), \ldots, \varrho_N(k)\}$.

- $1 > \varrho_1(k) \geq \rho_2(k) \geq \varrho_i(k) \geq \varrho_N(k) \geq 0$ are the **partial canonical correlations** between Y_t and Y_{t-k}.

Partial Cross-Correlations

Regularized Multivariate Durbin-Levinson Algorithm

- **Step 1: regularization** of the PCC matrices: replace $\mathcal{P}_k$ by

$$P_k = w(k)\text{diag}\{s_\tau(\varrho_1(k)), \ldots, s_\tau(\varrho_N(k))\},$$

where $w(k)$ are the weights of the trapezoidal kernel (see below) and $s_\tau(\cdot)$ is the Smoothly Clipped Absolute Deviation (SCAD) thresholding function (Fan and Li, 2001). Denote the regularized PCC matrix by $C(k) = L_k P_k R_k'$.

- **Step 2: reconstruction**. We invert the MDL algorithm to recover the coefficient matrices and the prediction error variance matrix of the regularized predictor:

$$\{\Pi_{m1}, \Pi_{m2}, \ldots, \Pi_{mm}; V_m, m = 1, \ldots, n\},$$

$$\tilde{Y}_{m,t} = \Pi_{m1} Y_{t-1} + \Pi_{m2} Y_{t-2} + \cdots + \Pi_{mm} Y_{t-m}.$$

- The solution preserves stationarity and positive definiteness of the cross-covariance function (spectral density matrix).

Reconstruction Details

Tapering and Thresholding

- The lags are regularized by the weighting function $w(k) = \kappa(k/\ell)$, where ℓ is the banding parameter and $\kappa(u)$ is the trapezoidal kernel (Politis and Romano, 1995)

$$\kappa(u) = \begin{cases} 1, & |u| \leq 1 \\ 2 - |u|, & 1 < |u| \leq 2 \\ 0, & |u| > 2 \end{cases}$$

- The SCAD penalty (Fan and Li 2001) regularizes the canonical correlations as follows (for $0 \leq \rho < 1$)

$$s_\tau(\rho) = \begin{cases} \max(0, \rho - \tau), & \rho \leq 2\tau, \\ \{(a-1)\rho - a\tau\}/(a-2), & 2\tau < \rho \leq a\tau, \\ \rho, & \rho > a\tau, \end{cases}$$

where $a = 3.7$ and $\tau \in (0, 1)$ is the threshold parameter.

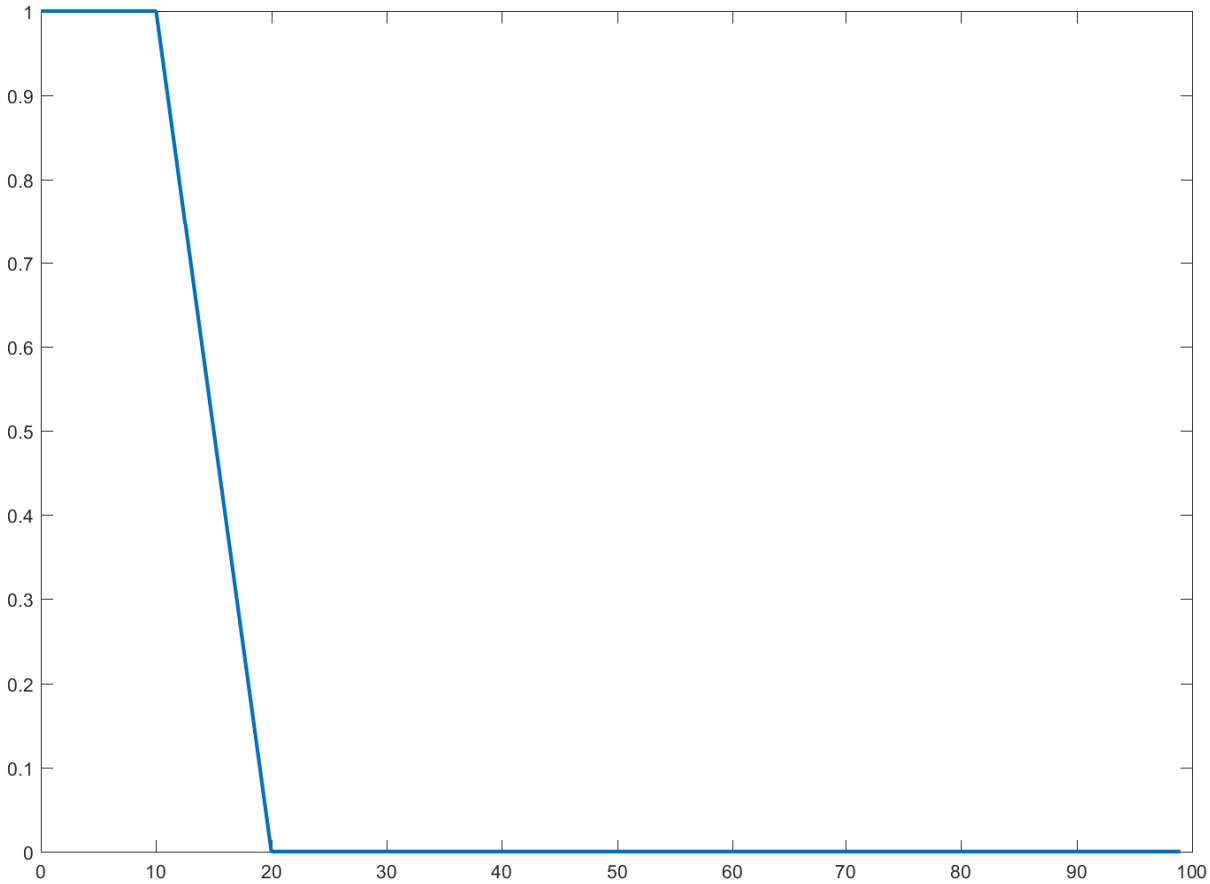


Figure: Trapezoidal kernel with $\ell = 10$ and $n = 100$

- Let $\{y_t, t = 1, \ldots, n\}$ denote the observed time series.
- If we assume that N is fixed, we estimate $\{\Pi_{mj}, j = 1, \ldots, m, V_m\}$ by running the regularized MDL on the sample cross-covariance matrices $\hat{\Gamma}(k) = n^{-1} \sum_{t=k+1}^{n} y_t y_{t-k}', k = 1, \ldots, m$.
- Our regularized MDL estimator receives as an input the sample cross-covariances and performs a dynamic shrinkage along the lag dimension.
- If N is large, relative to n, we need to further perform a cross-sectional shrinkage towards a finite order seemingly unrelated autoregressive system.

Regularizing along the cross-sectional dimension

- For high dimensional systems we apply the regularized MDL algorithm to the mixture sample cross-covariance (CC) matrix $\hat{\Gamma}_{\lambda,m} = \{\hat{\Gamma}_\lambda(|i-j|), i,j = 1,\ldots,m\}$, where, for $k = 0, 1, \ldots, m$,

$$\hat{\Gamma}_\lambda(k) = \lambda\hat{\Gamma}(k) + (1-\lambda)\hat{D}(k), \quad \hat{D}(k) = \text{diag}\{\hat{\gamma}_{ii}(k)\} \tag{3}$$

- $\hat{\Gamma}_\lambda(k)$ is a convex combination of the sample cross-covariance matrix at lag k and the matrix containing the sample autocovariances along the diagonal and the cross-covariances are set to zero.
- The weight $\lambda \in [0,1]$ regulates the shrinkage intensity towards the target cross-covariance matrix $\hat{D}(k)$.
- The resulting regularized MDL estimator shrinks the partial cross-correlation matrices towards zero along the lag dimension and along the cross-sectional dimension.

Statistical Inference

- The estimator depends on three key parameters:
  - The shrinkage intensity λ
  - The banding parameter ℓ of the trapezoidal kernel.
  - The threshold parameter τ of the SCAD penalty function.
- The Whittle algorithm receives as an input the mixture sample CC matrices $\hat{\Gamma}_\lambda(k) = \lambda\hat{\Gamma}(k) + (1-\lambda)\hat{D}(k)$, where $\hat{D}(k) = \text{diag}\{\hat{\gamma}_{ii}(k)\}$ and $\lambda \in [0,1]$.
- For $m = 1, \ldots, n-1$, the Regularized MDL algorithm returns the coefficient matrices $\tilde{\Pi}_{mj}, j = 1, \ldots, m$, and the prediction error covariance matrices $\tilde{V}_m$, as well as the regularized sample CC matrix $\tilde{\Gamma}_\lambda(k)$.
- The parameters λ, ℓ, τ are estimated by time series cross-validation.

Estimation Results

- A three-dimensional grid in $[0,1] \times [0.5,6] \times [0,0.4]$ is considered for the triple (λ, ℓ, τ), with steps (0.05, 0.25, 0.01).
- The size of the rolling window (training sample) is $T = 120$ (ten years of monthly observations).
- Let i be an index taking values from 1 to $n - T$.
- Using observations from time i to time $T + i - 1$ as a training sample, for each combination of values for (λ, ℓ, τ)
  - we construct the mixture sample cross-covariance (CC) matrix $\hat{\Gamma}_{\lambda,n}$
  - Obtain the coefficients of the linear predictor $\tilde{\Pi}_{mj}, j = 1, \ldots, m$, $m = \lfloor 2\ell \rfloor$. (These depend on τ too)
  - Predict recursively observations at times $T + i$ to $T + i + H$, where $H = 18$ (indirect predictor). Let $\hat{y}_{T+i+h-1|T+i-1}$ denote the h-step-ahead prediction.
  - Compare $y_{T+i+h-1}$ with $\hat{y}_{T+i+h-1|T+i-1}$.
- The experiment yields $n - T - h = 401 - h$ h-step-ahead prediction errors. The MSFE matrix for the target series is computed for horizons $h = 1, \ldots, 12$.

- Let $\hat{\Omega}_i = \{\hat{\omega}_{hk}, h, k = 1, \ldots, H\}$, denote the multistep prediction error covariance matrix for the i-th target series, where

$$\hat{\omega}_{hk} = \frac{1}{n^*} \sum_{j=1}^{n^*} (y_{i,T+j} - \hat{y}_{i,T+j|T+j-h})(y_{i,T+j} - \hat{y}_{i,T+j|T+j-k}),$$

  and $n^* = n - T - h \vee k$.

- The optimal values of (λ, ℓ, τ) are the minimizers of

$$\hat{\varsigma}_{tr} = \text{trace}(\hat{\Omega}_i)/H,$$

  or

$$\hat{\varsigma}_{det} = |\hat{\Omega}_i|^{1/H}.$$

Table: Niño 3.4: selection of (λ, ℓ, τ) values for $H = 12$.

Trace				Determinant			
$\hat{\lambda}$	$\hat{\ell}$	$\hat{\tau}$	$\hat{\varsigma}_{tr}$	$\hat{\lambda}$	$\hat{\ell}$	$\hat{\tau}$	$\hat{\varsigma}_{det}$
0.50	1.25	0.04	0.49	0.50	1.00	0.00	0.0398
0.55	**1.50**	**0.06**	**0.48**	0.55	1.00	0.00	0.0398
0.60	1.50	0.09	0.49	0.60	1.00	0.01	0.0399
0.65	1.50	0.03	0.50	0.65	1.00	0.05	0.0399
0.70	1.50	0.00	0.49	0.70	1.00	0.06	0.0400
0.75	1.50	0.09	0.50	0.75	1.00	0.06	0.0400
0.80	1.50	0.10	0.50	0.80	1.00	0.02	0.0401
0.85	2.75	0.32	0.51	0.85	1.00	0.02	0.0403
0.90	1.50	0.15	0.52	0.90	1.00	0.04	0.0405
0.95	1.50	0.02	0.52	0.95	1.00	0.02	0.0407

Table: SOI: selection of (λ, ℓ, τ) values for $H = 12$.

Trace				Determinant			
$\hat{\lambda}$	$\hat{\ell}$	$\hat{\tau}$	$\hat{\varsigma}_{tr}$	$\hat{\lambda}$	$\hat{\ell}$	$\hat{\tau}$	$\hat{\varsigma}_{det}$
0.50	1.25	0.02	1.91	0.50	1.25	0.03	1.3413
0.55	**1.50**	**0.02**	**1.90**	0.55	1.25	0.00	1.3423
0.60	1.25	0.15	1.92	0.60	1.00	0.27	1.3436
0.65	1.50	0.13	1.92	0.65	1.25	0.26	1.3444
0.70	1.25	0.00	1.93	0.70	1.25	0.00	1.3429
0.75	1.50	0.19	1.96	0.75	1.25	0.11	1.3488
0.80	1.50	0.14	1.96	0.80	1.00	0.13	1.3514
0.85	1.50	0.23	1.98	0.85	1.25	0.10	1.3531
0.90	1.50	0.30	2.00	0.90	1.25	0.30	1.3577
0.95	1.25	0.05	2.02	0.95	1.25	0.37	1.3594

Comparison with VAR model

- The selection of the VAR order follows the same criteria (based on out-of-sample performance).

	Niño 3.4			SOI	
Order	$\hat{\varsigma}_{tr}$	$\hat{\varsigma}_{det}$	Order	$\hat{\varsigma}_{tr}$	$\hat{\varsigma}_{det}$
1	0.7121	0.0000	1	0.0005	2.2558
2	0.6286	0.0006	2	0.0143	2.3550
3	0.7277	0.0040	3	0.1382	2.8714
4	1.0868	0.0135	4	0.4405	3.5837
5	1.4947	0.0358	5	1.1665	5.2725

- We take $p = 1$.

Table: Relative Root Mean Square Error of rMDL versus VAR(1) model forecasts

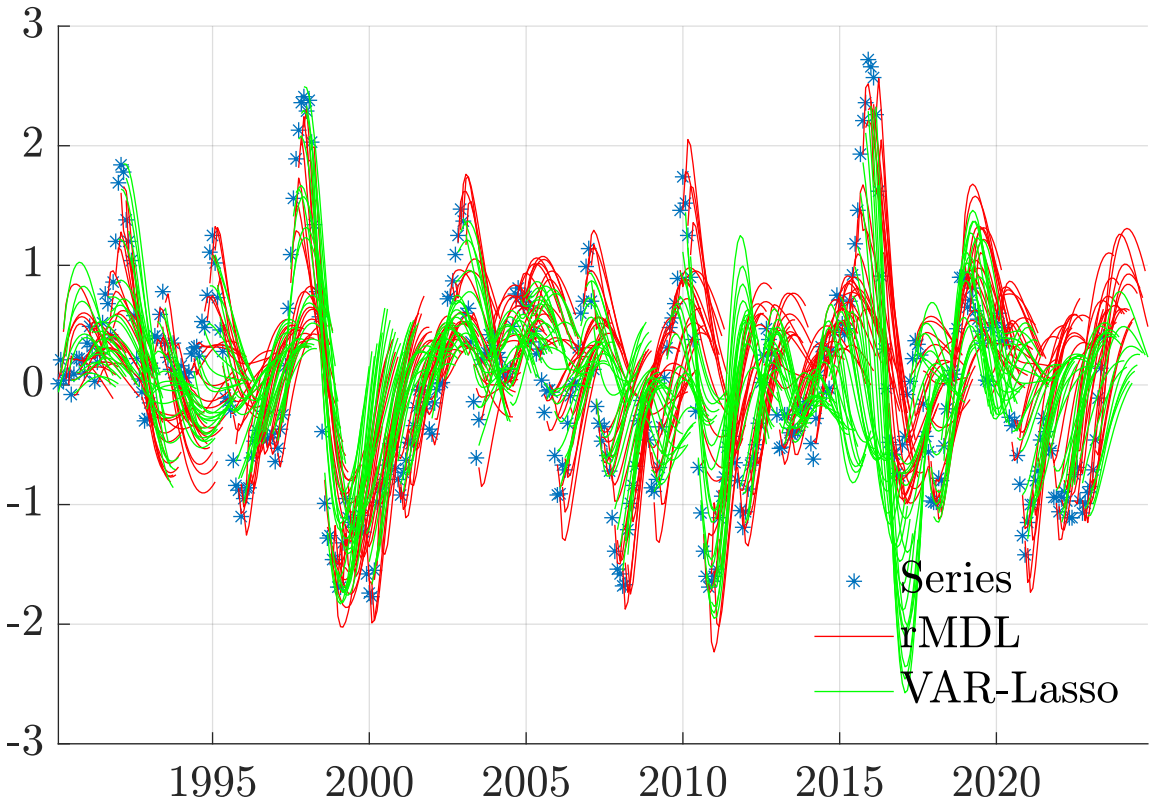
	Forecast horizon					
Series	1	2	3	4	5	6
Niño 1+2	1.006	1.260	1.255	1.183	1.116	1.058
Niño 3	1.116	1.334	1.182	1.064	0.985	0.932
Niño 4	1.004	1.180	1.077	0.993	0.935	0.896
Niño 3.4	1.088	1.248	1.073	0.971	0.908	0.869
SLP Tahiti	0.942	0.998	0.993	0.986	0.977	0.963
SLP Darwin	0.964	1.019	1.002	1.002	0.997	0.985
Series	7	8	9	10	11	12
Niño 1+2	1.012	0.968	0.935	0.913	0.893	0.875
Niño 3	0.887	0.846	0.817	0.797	0.779	0.765
Niño 4	0.870	0.850	0.835	0.823	0.816	0.813
Niño 3.4	0.839	0.812	0.791	0.775	0.763	0.753
SLP Tahiti	0.945	0.921	0.907	0.892	0.884	0.875
SLP Darwin	0.964	0.953	0.939	0.921	0.906	0.894
Series	13	14	15	16	17	18
Niño 1+2	0.859	0.848	0.839	0.831	0.824	0.817
Niño 3	0.753	0.745	0.740	0.737	0.738	0.740
Niño 4	0.811	0.813	0.818	0.825	0.834	0.845
Niño 3.4	0.745	0.742	0.742	0.743	0.747	0.752
SLP Tahiti	0.873	0.874	0.876	0.879	0.885	0.891
SLP Darwin	0.886	0.881	0.875	0.871	0.869	0.863

Comparison with the VAR-Lasso predictor and combination

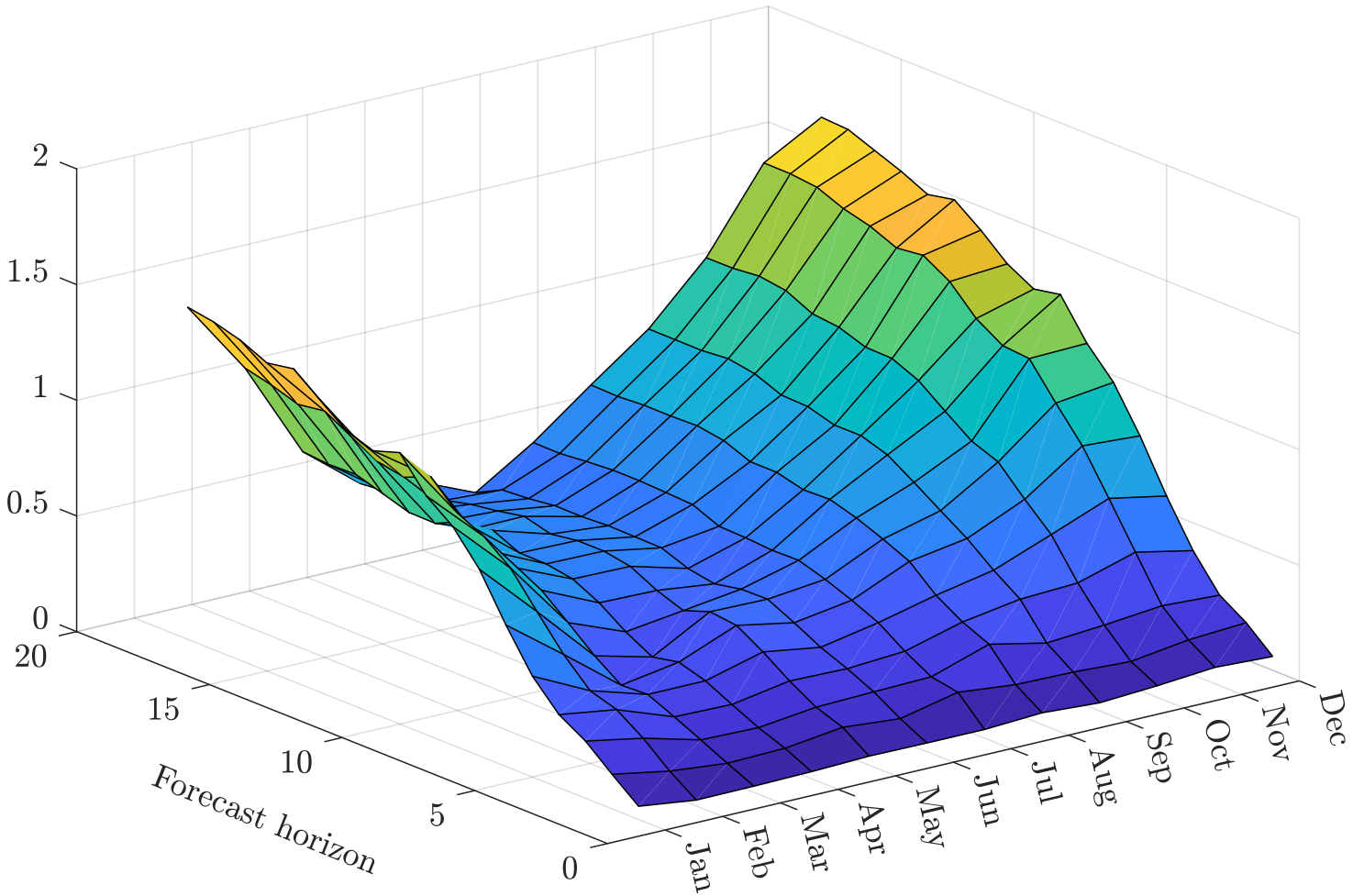
Table: Mean Square Forecast Error comparison

	Forecast horizon					
Predictor	1	2	3	4	5	6
rMDL	0.043	0.120	0.196	0.288	0.385	0.484
VAR-Lasso	0.037	0.059	0.116	0.203	0.313	0.439
Combined	0.037	0.062	0.093	0.141	0.197	0.258
	Forecast horizon					
Predictor	7	8	9	10	11	12
rMDL	0.574	0.646	0.706	0.753	0.789	0.816
VAR-Lasso	0.573	0.706	0.833	0.948	1.047	1.129
Combined	0.317	0.372	0.423	0.469	0.512	0.552
	Forecast horizon					
Predictor	13	14	15	16	17	18
rMDL	0.835	0.854	0.869	0.880	0.889	0.893
VAR-Lasso	1.191	1.235	1.262	1.274	1.273	1.261
Combined	0.590	0.621	0.644	0.656	0.660	0.654

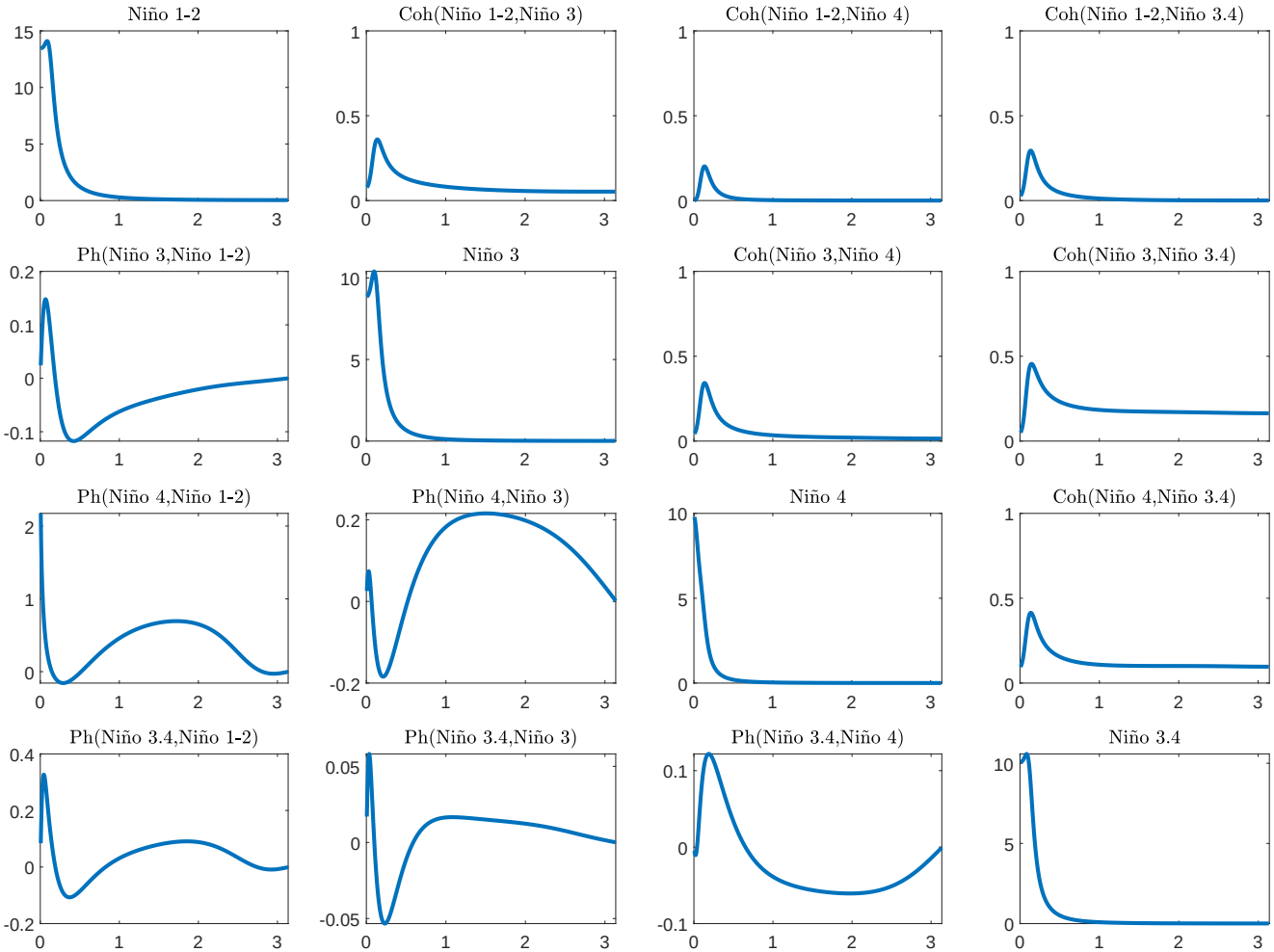
Rolling forecasts vintages $H = 18$ for $\hat{\lambda} = 0.55, \hat{\ell} = 1.50, \hat{\tau} = 0.06$.



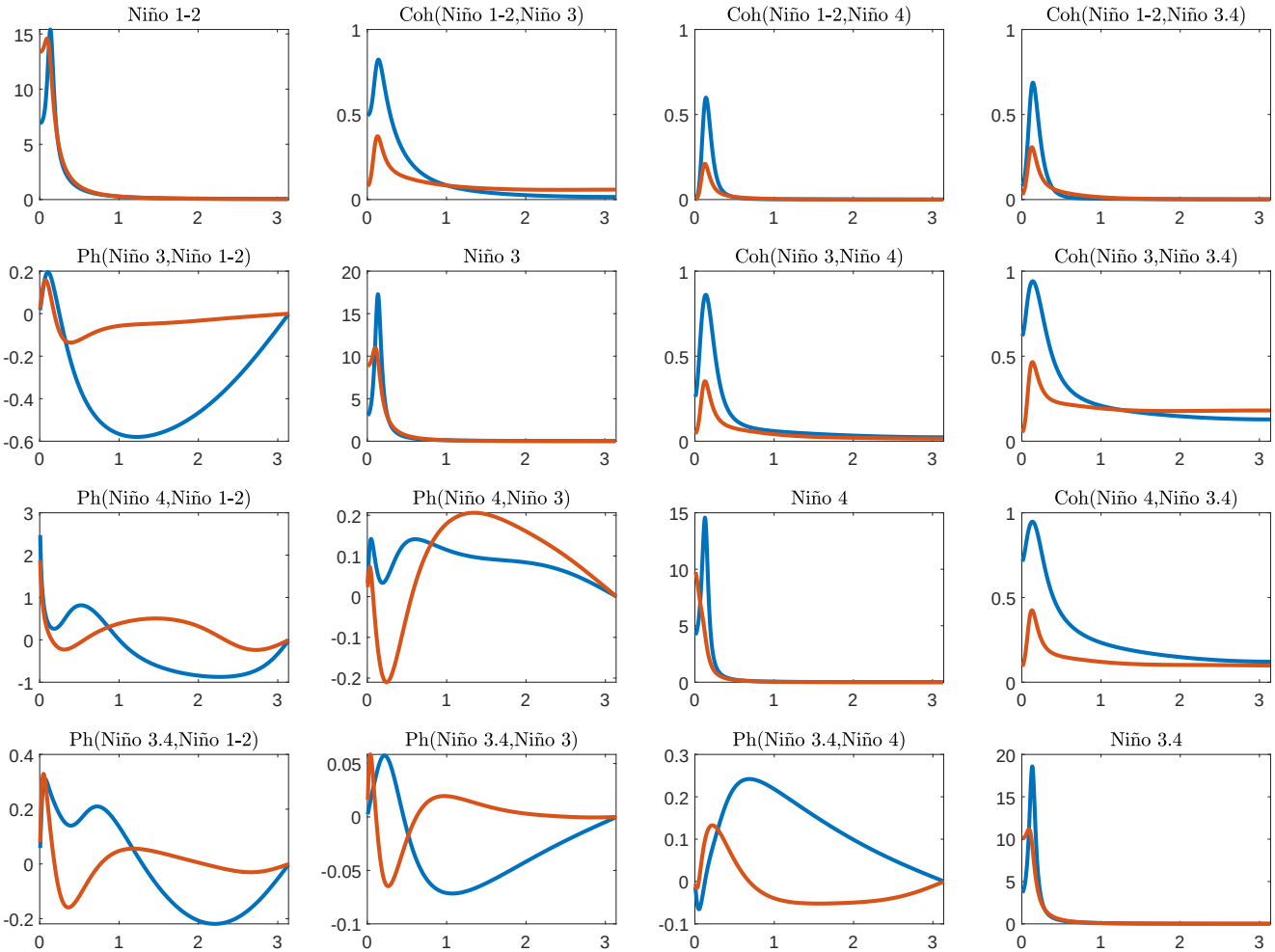
Mean square forecast error by month and forecast horizon for Niño 3.4. The plot is a confirmation of the spring predictability barrier: ENSO predictability is lower during the boreal spring.



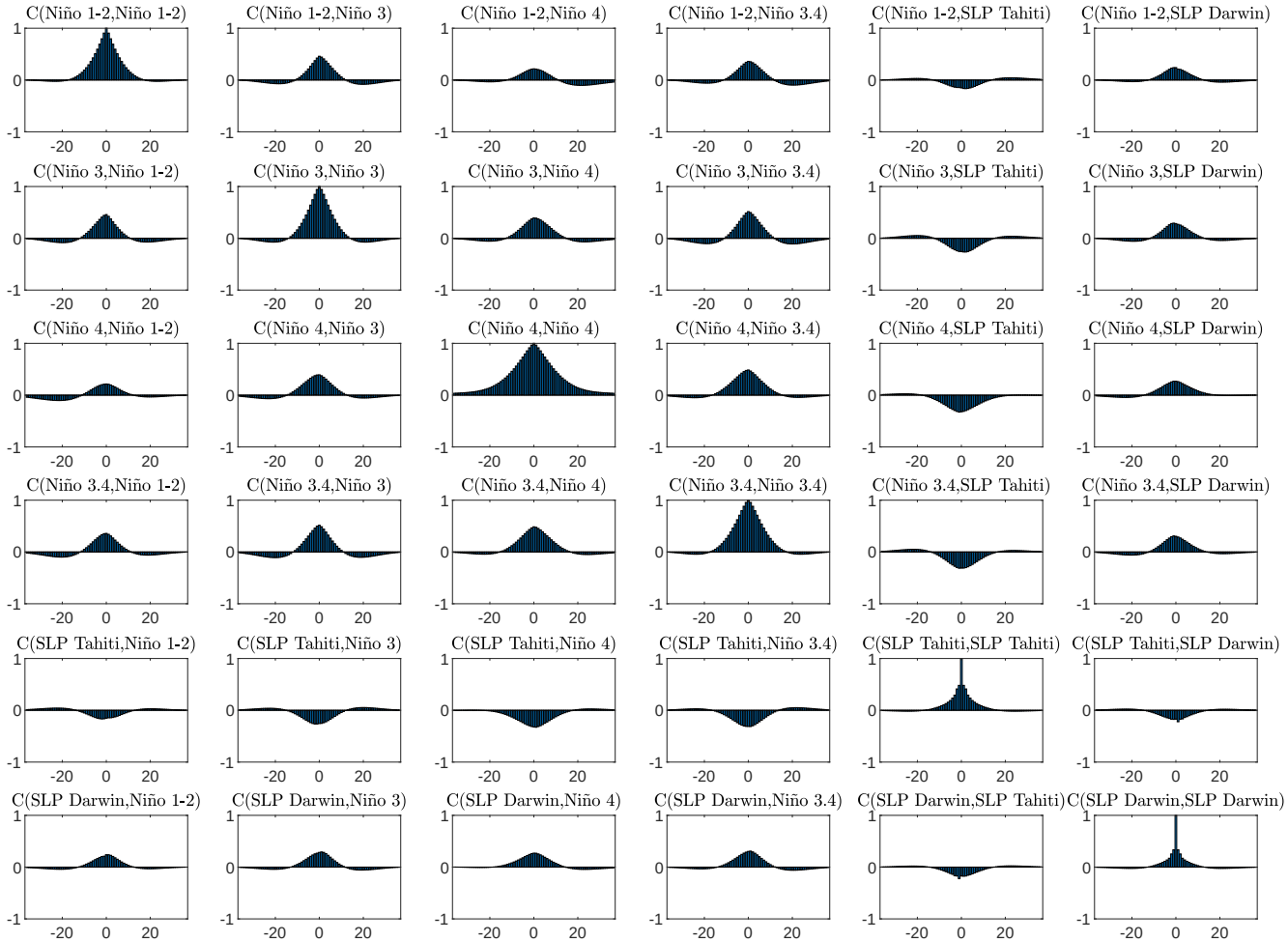
Spectral density estimates ($\hat{\lambda} = 0.55, \hat{\ell} = 1.50, \hat{\tau} = 0.06$)



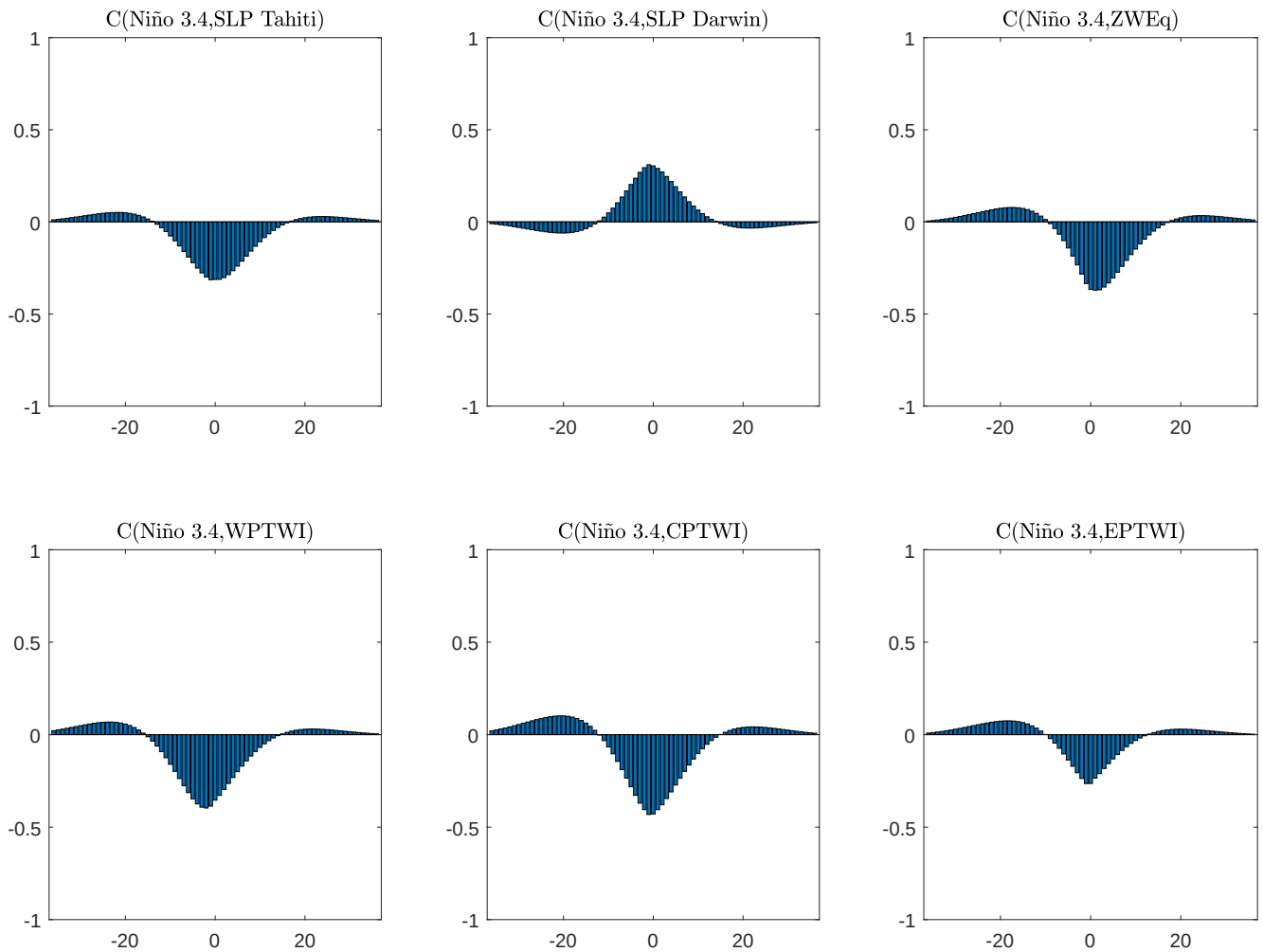
Comparison with VAR-LASSO estimated spectra (blue) and rMDL (red).



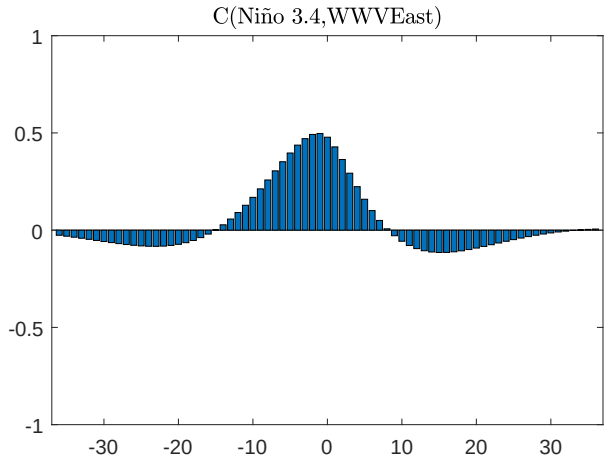
Cross correlations of SST series

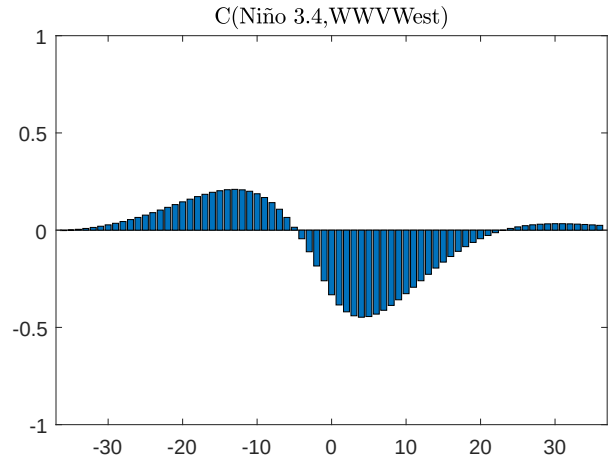


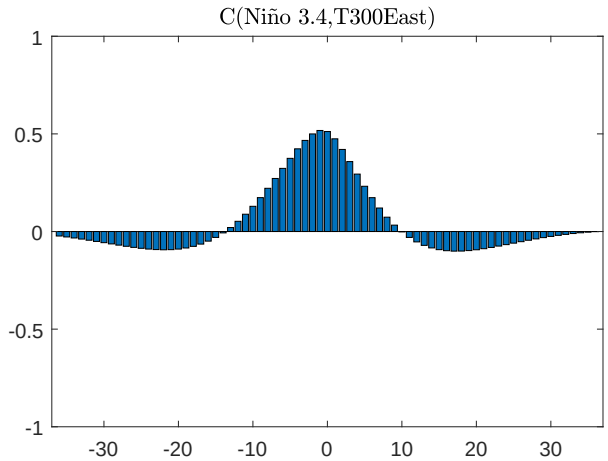
Cross correlations between Niño 3.4 and SLP, and Trade Winds series.

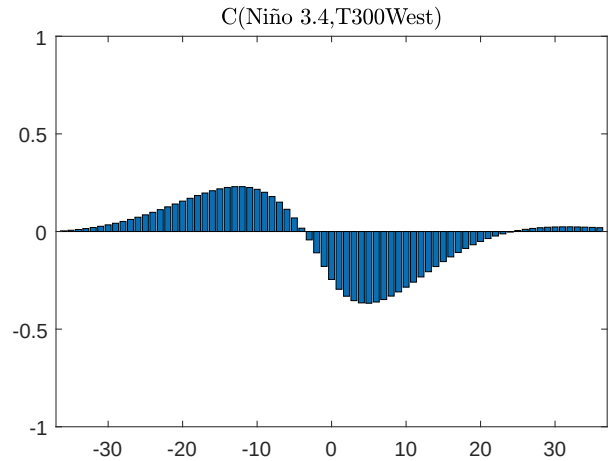


Cross correlations between Nino 3-4 and Upper Ocean Heat content.







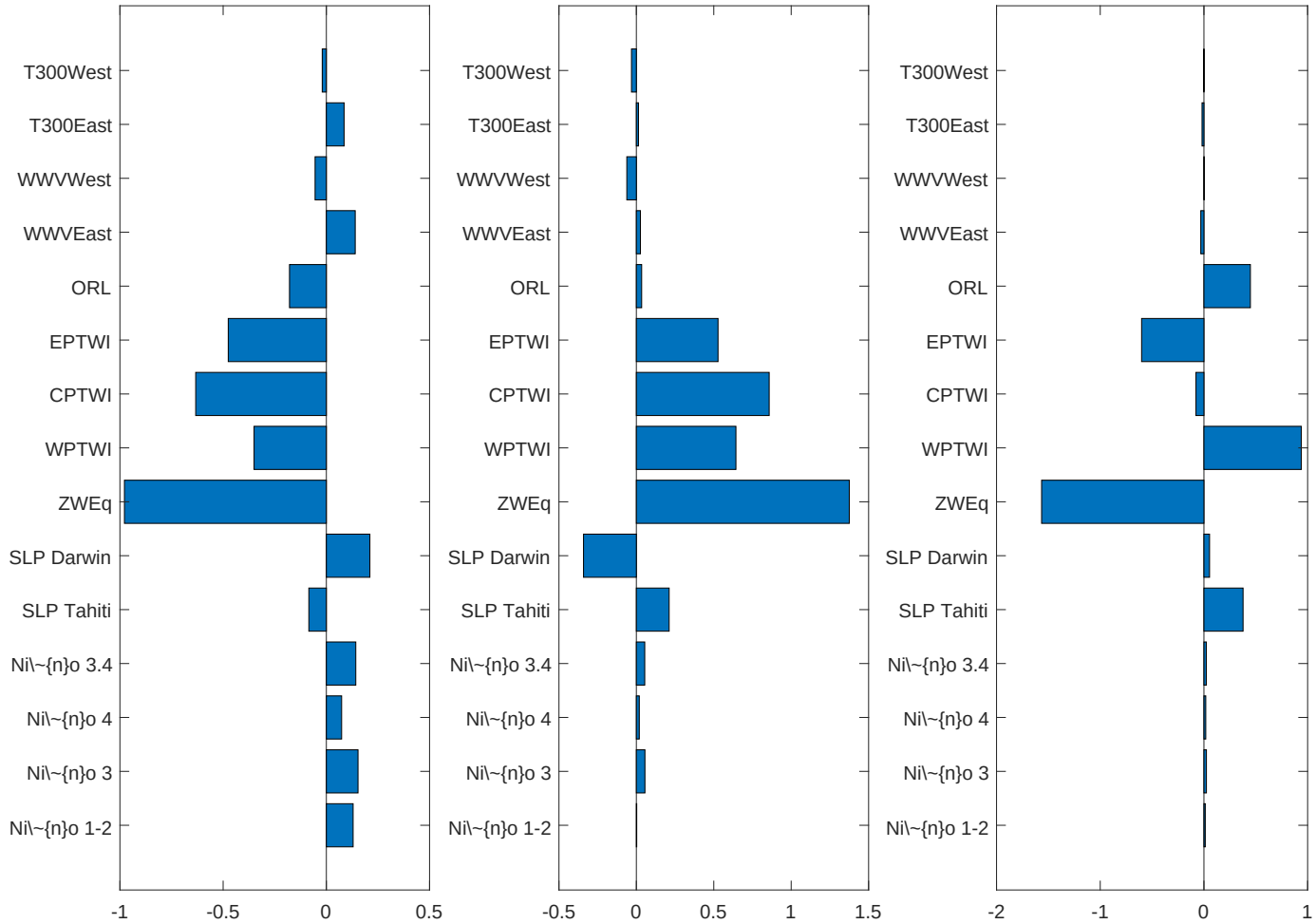


Structural Analysis

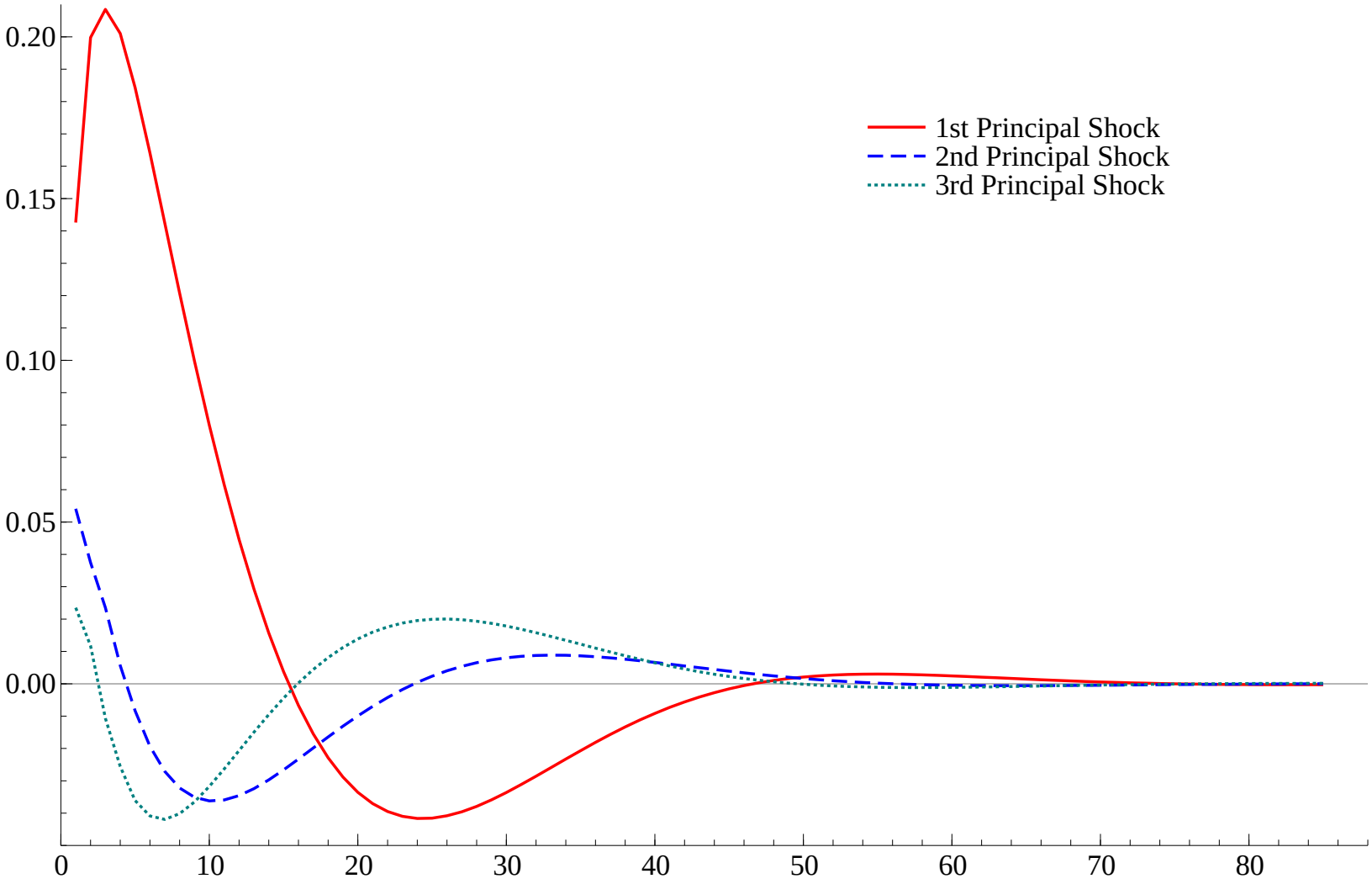
- Let U_t denotes the reduced form shocks, $U_t = Y_t - \sum_j \Pi_j Y_{t-j}$, and let $\text{Var}(U_t) = V$.
- Write $V = DCD$, where $D = \text{diag}(\sqrt{v_1}, \ldots, \sqrt{v_N})$, and $C = D^{-1}VD^{-1}$ is the correlation matrix of the reduced form shocks.
- Let $C = A\Lambda A'$ be the spectral decomposition of C.
- Define unit variance orthogonal (structural) shocks $\epsilon_t = \Theta_0^{-1} U_t$, where $\Theta_0 = DA\Lambda^{1/2}$ is such that $\Theta_0\Theta_0' = V$.
- Then, $U_t = \Theta_0 \epsilon_t$, and the structural impulse response function (IRF) is the sequence of coefficients associated to the powers of z in

$$(I_N - \sum_j \Pi_j z^j)^{-1}\Theta_0.$$

Loadings of U_t on ϵ_t.



Impulse response functions of series Niño 3.4 to first three principal shocks.



Conclusive remarks

We have proposed a(n auto-) cross-covariance estimator based on the idea of regularizing the sample partial cross-correlations. The estimator has the following comparative advantages:

- The estimated cross-covariance function is positive definite by construction and no positivity correction is needed.
- The optimal linear predictor based on a sample of size n requires a number of operations which is $O(N^2\ell^2)$.
- It performs well in the empirical application, bringing out the cyclical features of the multivariate time series.

Multivariate Durbin-Levinson

- Initialization.

$$\begin{aligned} \Sigma_0 &= \Gamma(0), & \Sigma_0^* &= \Gamma(0), \\ \Delta_0 &= \Gamma(1), & \Delta_0^* &= \Gamma(1)', \end{aligned}$$

- For $m = 1, 2 \ldots, n-1$,
  - compute the partial autoregressive matrices

$$\Phi_{mm} = \Delta_{m-1}\Sigma_{m-1}^{*-1}, \qquad \Phi_{mm}^* = \Delta_{m-1}^*\Sigma_{m-1}^{-1},$$

  - If $m > 1$, for $k = 2, \ldots, m-1$,

$$\Phi_{mk} = \Phi_{m-1,k} - \Phi_{mm}\Phi_{m-1,m-k}^*, \quad \Phi_{mk}^* = \Phi_{m-1,k}^* - \Phi_{mm}^*\Phi_{m-1,m-k},$$

Evaluate

$$\Sigma_m = \Gamma(0) - \sum_{k=1}^m \Phi_{mk}\Gamma(-k), \qquad \Sigma_m^* = \Gamma(0) - \sum_{k=1}^m \Phi_{mk}^*\Gamma(k)$$

$$\begin{aligned} \Delta_m &= \Gamma(m+1) - \sum_{k=1}^m \Phi_{mk}\Gamma(m-k+1), \\ \Delta_m^* &= \Gamma(-m-1) - \sum_{k=1}^m \Phi_{mk}^*\Gamma(k-m-1). \end{aligned}$$

- The matrices $\{\Phi_{mk}, k = 1, \ldots, m\}$ are the coefficients matrices of the minimum mean square linear forward predictor based on m past observations, $\hat{Y}_{mt} = \Phi_{m1} Y_{t-1} + \Phi_{m2} Y_{t-2} + \cdots + \Phi_{mm} Y_{t-m}$.
  They arise as the solution of the Yule-Walker matrix equations $[\Phi_{m1}, \Phi_{m2}, \ldots, \Phi_{mm}] = [\Gamma(1), \Gamma(2), \ldots, \Gamma(m)]\, \Gamma_m^{-1}$.
- The matrices $\{\Phi^*_{mk}, k = 1, \ldots, m\}$ are the coefficients matrices of the minimum mean square linear backward predictor based on m future observations, $\hat{Y}^*_{m,t-m} = \Phi^*_{m1} Y_{t-m+1} + \Phi^*_{m2} Y_{t-m+2} + \cdots + \Phi^*_{mm} Y_t$, and are the solution of the Yule-Walker matrix equations $[\Phi^*_{m1}, \Phi^*_{m2}, \ldots, \Phi^*_{mm}] = [\Gamma(-1), \Gamma(-2), \ldots, \Gamma(-m)]\, \Gamma_m^{-1}$.
- The matrices Σ_m and Σ^*_m are the variance-covariance matrices of the forward and backward prediction errors $\mathcal{U}_{m,t} = Y_t - \hat{Y}_{m,t}$, and $\mathcal{U}^*_{m,t-m} = Y_{t-m} - \hat{Y}^*_{m,t-m}$, respectively. Finally, $\Delta_{m-1} = \mathrm{Cov}(\mathcal{U}_{m-1,t}, \mathcal{U}^*_{m-1,t-m})$ and $\Delta^*_{m-1} = \mathrm{Cov}(\mathcal{U}^*_{m-1,t-m}, \mathcal{U}_{m-1,t})$.
  Obviously, $\Delta^{'}_{m-1} = \Delta^*_{m-1}$.

Back

Partial Cross-Correlation Matrices

Writing $\Sigma_{m-1} = \Sigma_{m-1}^{1/2}\Sigma_{m-1}^{1/2'}$ and $\Sigma_{m-1}^{*} = \Sigma_{m-1}^{*1/2}\Sigma_{m-1}^{*1/2'}$, the partial cross-correlation matrix at lag m,

$$\mathcal{C}(m) = \Sigma_{m-1}^{-1/2}\Phi_{mm}\Sigma_{m-1}^{*1/2},$$

is the correlation matrix of the 'standardized' forward and backward prediction errors $\mathcal{E}_{m-1,t} = \Sigma_{m-1}^{-1/2}\mathcal{U}_{m-1,t}$ and $\mathcal{E}_{m-1,t-m}^{*} = \Sigma_{m-1}^{*-1/2}\mathcal{U}_{m-1,t-m}^{*}$. See, e.g., Morf et al. (1978) and Reinsel (2003).
The squared partial canonical correlations between $\mathcal{U}_{m-1,t}$ and $\mathcal{U}_{m-1,t-m}^{*}$, $\varrho_i^2(m), i = 1, \ldots, N$, are the eigenvalues of $\mathcal{C}(m)\mathcal{C}(m)'$. The matrices $\Phi_{mm}\Phi_{mm}^{*}$ and $\mathcal{C}(m)\mathcal{C}(m)'$ are similar and thus have the same eigenvalues.

Back

Reconstruction

The regularized partial cross-correlation matrix is

$$C(m) \quad = \quad L_m P(m) R'_m,$$

$P(m) = \text{diag}\{\rho_i(m), i = 1, \ldots, N\}$, $\rho_i(m) = w(m) s_\tau(\varrho_i(m))$.
From $\Gamma(0)$ and $\{C(k), k = 1, \ldots, m\}$ we reconstruct the regularized coefficient matrices $\{\Pi_{mj}, j = 1, \ldots, m\}$, of the forward linear predictor of Y_t based on m past observations $\{Y_{t-j}, j = 1, \ldots, m\}$:

$$\tilde{Y}_{m,t} = \Pi_{m1} Y_{t-1} + \Pi_{m2} Y_{t-2} + \cdots + \Pi_{mm} Y_{t-m}, \tag{4}$$

and the regularized cross-covariance at lag m,

$$\Gamma_r(m) = \Pi_{m-1,1} \Gamma_r(m-1) + \cdots + \Pi_{m-1,m-1} \Gamma_r(1) + \Pi_{mm} V_{m-1},$$

($\Gamma_r(0) = \Gamma(0)$). Moreover, it delivers the regularized forward prediction error covariance matrix V_m.

- Initialization.

$$V_0 = \Gamma(0), \qquad V_0^* = \Gamma(0),$$

- For $m = 1, 2 \ldots, n-1$,
  - Factorize the matrices

$$V_{m-1} = V_{m-1}^{1/2} V_{m-1}^{1/2'}, \qquad V_{m-1}^* = V_{m-1}^{*1/2} V_{m-1}^{*1/2'},$$

  - compute the regularized partial autoregressive matrices

$$\Pi_{mm} = V_{m-1}^{1/2} C(m) V_{m-1}^{*-1/2}, \qquad \Pi_{mm}^* = V_{m-1}^{*1/2} C(m)' V_{m-1}^{-1/2},$$

  - If $m > 1$, for $k = 2, \ldots, m-1$,

$$\Pi_{mk} = \Pi_{m-1,k} - \Pi_{mm}\Pi_{m-1,m-k}^*, \quad \Pi_{mk}^* = \Pi_{m-1,k}^* - \Pi_{mm}^*\Pi_{m-1,m-k},$$

    Evaluate

$$V_m = (I_N - \Pi_{mm}\Pi_{mm}')V_m, \qquad V_m^* = (I_N - \Pi_{mm}'\Pi_{mm})V_m^*$$

- the regularized cross-covariance at lag m, are obtained recursively as

$$\Gamma_r(m) = \Pi_{m-1,1}\Gamma_r(m-1) + \cdots + \Pi_{m-1,m-1}\Gamma_r(1) + \Pi_{mm}V_{m-1}, \tag{5}$$

  where $\Gamma_r(0) = \Gamma(0)$.
- The matrices V_m and V_m^* are the variance-covariance matrices of the forward and backward prediction errors $U_{m,t} = Y_t - \tilde{Y}_{m,t}$, and $U^*_{m,t-m} = Y_{t-m} - \tilde{Y}^*_{m,t-m}$, respectively, where $\tilde{Y}_{m,t} = \Pi_{m1}Y_{t-1} + \Pi_{m2}Y_{t-2} + \cdots + \Pi_{mm}Y_{t-m}$ is the regularized forward predictor and $\tilde{Y}^*_{m,t-m} = \Pi^*_{m1}Y_{t-m+1} + \Pi^*_{m2}Y_{t-m+2} + \cdots + \Pi^*_{m,m}Y_t$ is the regularized backward predictor.

Back