We consider two general approaches to building spatio-temporal models:

- Models that specify a space-time correlation function.
- Models based on linear representations of the process that assume that some of the linear components are time-varying.

We consider two general approaches to building spatio-temporal models:

- Models that specify a space-time correlation function.
- Models based on linear representations of the process that assume that some of the linear components are time-varying.

We consider Gaussian space-time processes denoted as $X(s,t)$ where $(s,t) \in S \times \mathbb{R}^+$. Usually S will correspond to $\mathbb{R}^2$ or $\mathbb{R}^3$ and t will be measured at discrete times. We denote the covariance function as $C(s,s',t,t')$

The full specification of the covariance of a process observed at times $t = 1, \ldots, T$ and locations $s_1, \ldots, s_n$, requires $nT(nT + 1)/2$ parameters. Even for moderate T, this is can be a very large number.

The full specification of the covariance of a process observed at times $t = 1, \dots, T$ and locations $s_1, \dots, s_n$, requires $nT(nT+1)/2$ parameters. Even for moderate T, this is can be a very large number.

Assuming separability we have that $C(s, s', t, t') = C_1(s, s')C_2(t, t')$, so denoting Σ the covariance matrix, we have that $\Sigma = V \otimes W$ where $V_{i,j} = C_1(s_i, s_j)$ and $W_{k,l} = C_2(k, l)$, $V \in \mathbb{R}^{n \times n}$ and $W \in \mathbb{R}^{T \times T}$. So now there are only $(n(n+1) + T(T+1))/2$ parameters.

Separability is computationally convenient but it has strong implications on the model. If $\rho(s, s', t, t')$ is the correlation function, then

$$\rho(s, s', t, t') = \frac{C_1(s, s')C_2(t, t')}{\sqrt{C_1(s, s)C_1(s', s')C_2(t, t)C_2(t', t')}}$$

implies that $\rho(s, s', t, t) = \rho_1(s, s')$, $\forall t$ and $\rho(s, s, t, t') = \rho(t, t')$, $\forall s$.

Separability is computationally convenient but it has strong implications on the model. If $\rho(s, s', t, t')$ is the correlation function, then

$$\rho(s, s', t, t') = \frac{C_1(s, s')C_2(t, t')}{\sqrt{C_1(s, s)C_1(s', s')C_2(t, t)C_2(t', t')}}$$

implies that $\rho(s, s', t, t) = \rho_1(s, s'),\ \forall t$ and $\rho(s, s, t, t') = \rho(t, t'),\ \forall s$.

Thus the correlation between $X(s, t)$ and $X(s', t)$ does not depend on time, and the correlation between $X(s, t)$ and $X(s, t')$ does not depend on the site. This is an undesirable property, since many natural phenomena have time-varying spatial correlations.

Non-Separable Correlations

By modeling the spectral densities of stationary processes we can obtain correlation functions that are not separable. Let $h(\omega, v) = \rho(\omega, v)k(\omega)$, where ω denotes the spatial frequency and v the temporal one. The inverse Fourier transform of h will produce non-separable correlations.

By modeling the spectral densities of stationary processes we can obtain correlation functions that are not separable. Let $h(\omega, v) = \rho(\omega, v)k(\omega)$, where ω denotes the spatial frequency and v the temporal one. The inverse Fourier transform of h will produce non-separable correlations.

A class of covariance functions that are smooth everywhere, except possibly at the origin is given by the spectra:

$$f(\omega, v) = (c_1(a_1^2 + |\omega|^2)^{\alpha_1} + c_2(a_2^2 + v^2)^{\alpha_2})^{-\nu}.$$

This class allows for different degrees of smoothness across space than across time. It can also model long range dependencies.

A class that provides a direct construction of valid covariance functions, without the need of inverting the spectral density is

$$C(u,h) = \frac{\sigma^2}{\psi(|h|^2)^{d/2}}\,\phi\left(\frac{||u||^2}{\psi(|h|^2)}\right),\ (u,h) \in \mathbb{R}^d \times \mathbb{R}$$

where ϕ is a *completely monotone function*, i.e.,

$$\phi(t) = \int_0^\infty \exp(-rt)dF(r)$$

for some nondecreasing F. ψ is a positive function with completely monotone derivative.

Computational Issues

The above mentioned approaches to obtain wide classes of covariance functions have several computational problems.

- Inversion of the spectral density is possible only in a limited number of cases.

- For regularly spaced locations, it is possible to use spectral methods and fast approximations to the inverse Fourier transform.

- Estimating the parameters of a joint space and time covariance using likelihood based methods involves the computation of very large matrices.

Dynamic Conditionally Linear Models

The application of state space models to spatio-temporal modeling provides a framework that can accommodate time-varying dependencies and is suitable for handling large date sets, both in time and space.

Dynamic conditionally linear models (DCLM) can be written as

$$\begin{aligned} \boldsymbol{Y}_t &= F_t'\theta_t + \varepsilon_t, & \varepsilon &\sim N(0, V_t) \\ \theta_t &= G_t\theta_{t-1} + \omega_t, & \omega_t &\sim N(0, W_t), \end{aligned}$$

where $\boldsymbol{Y}_t$ are the observations at time t and θ_t is the state vector. F_t, G_t, V_t and W_t define the model and can be dependent on a set of parameters.

Dynamic Conditionally Linear Models

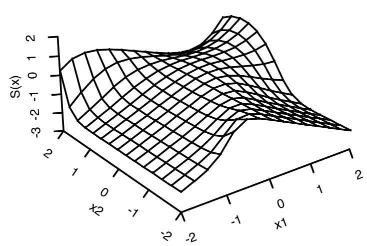
For location s assume a spatial model of the form

$$Y(s) = S(s; \boldsymbol{\beta}) + \varepsilon(s), \quad S(s; \boldsymbol{\beta}) = \sum_{j=1}^{J} \pi_j(s) \boldsymbol{f}_j'(s) \boldsymbol{\beta}_j$$

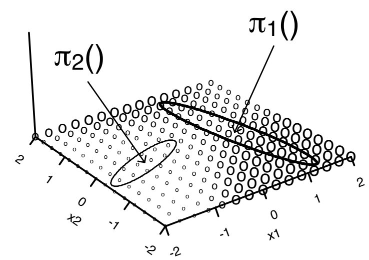
where $\boldsymbol{f}_j(s) = \{f_{j1}(s), \ldots, f_{jq}(s)\}'$ is a set of known basis functions. $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jq})'$ is a vector of parameters, $\pi_j(s)$ is a non-negative weighting kernel centered around location μ_j.

Dynamic Conditionally Linear Models

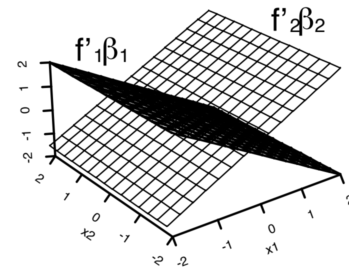
Spatial Process



Weighting Kernels



Component Surfaces

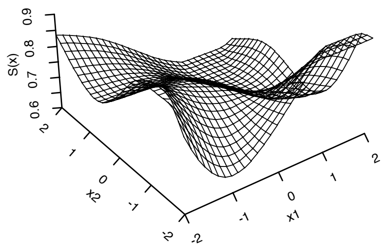


$$S(s;\boldsymbol{\beta}) \quad = \quad \sum_{j=1}^{2} \pi_j(s) \quad \times \quad \boldsymbol{f}_j'(s)\beta_j$$

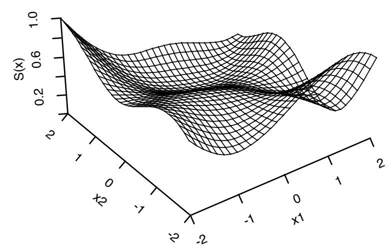
Spatial process generated from a mixture with $J = 2$. $\pi_j(\cdot)$ are Gaussian weighting kernels. $\boldsymbol{f}_j'(s)\boldsymbol{\beta}_j$ are linear regression surfaces.

Dynamic Conditionally Linear Models

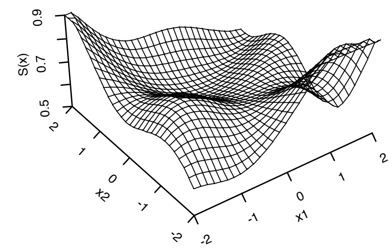
Time 1



Time 2



Time 3



$$S(s, \boldsymbol{\beta}_1) \longrightarrow S(s, \boldsymbol{\beta}_2) \longrightarrow S(s, \boldsymbol{\beta}_3)$$

A time series of spatial mean fields obtained through a locally-weighted mixture of $J = 6$ components with fixed weighting kernels and dynamic regression coefficients.

The estimation of the state parameters of a dynamic linear model corresponds to the sequential updating of a multivariate normal-normal model.

Posterior at $t-1$: $p(\theta_{t-1}|D_{t-1}) = N(m_{t-1}, C_{t-1})$

Prior at t: $p(\theta_t|D_{t-1}) = N(a_t, R_t)$

$$a_t = G_t m_{t-1} \ , \quad R_t = G_t C_{t-1} G_t' + W_t$$

One-step ahead forecast: $p(Y_t|D_{t-1}) = N(f_t, Q_t)$

Posterior at t: $p(\theta_t|D_t) = N(m_t, C_t)$

$$m_t = a_t + A_t e_t \ , \quad C_t = R_t - A_t Q_t A_t'$$

$$A_t = R_t F_t Q_t^{-1} \ , \quad e_t = Y_t - f_t$$

DLM Backward Equations

To sample the state vector in one block we use the smoothing equations. We start with a sample of

$$p(\theta_T|D_T) = N(m_T, C_T)$$

then sample

$$p(\theta_t|\theta_{t+1}, D_T) = N(h_t, H_t),\ t = T-1, \ldots, 1$$

where we calculate recursively

$$h_t = m_t + B_t(\theta_{t+1} - a_{t+1})$$

$$H_t = C_t - B_t R_{t+1} B_t' \ , \quad B_t = C_t G_{t+1}' R_{t+1}^{-1}.$$

This produces a sample of $p(\theta_1, \ldots, \theta_T|D_T)$.

To explore the posterior distribution of the state parameters in a DCLM we use forward filtering backwards sampling (FFBS).

The FFBS consists of running the Kalman forward equations until $t = T$, and then sample the state parameters at each step of the backwards smoother.

If the state parameters have a large dimension and T is large, FFBS can be very demanding computationally.

Blocking improves the mixing and reduces the computations.

Suppose that V_t and W_t have a common scale factor σ^2. Suppose that the model is a DLM conditional on a low dimensional parameter, say λ. Notice that $p(\boldsymbol{\beta}, \sigma^2, \lambda) = p(\boldsymbol{\beta}|\sigma^2, \lambda)p(\sigma^2|\lambda)p(\lambda)$.

Suppose that V_t and W_t have a common scale factor σ^2. Suppose that the model is a DLM conditional on a low dimensional parameter, say λ. Notice that $p(\boldsymbol{\beta}, \sigma^2, \lambda) = p(\boldsymbol{\beta}|\sigma^2, \lambda)p(\sigma^2|\lambda)p(\lambda)$.

From the theory of DLMs we have that

$$p(\lambda, \sigma^2|D_T) \propto p(\lambda)\, p(\sigma^2)$$
$$\prod_{t=1}^{T} |\boldsymbol{Q}_t|^{-1/2}(1/\sigma^2)^{T/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(\boldsymbol{Y}_t - \boldsymbol{f}_t)'\boldsymbol{Q}_t^{-1}(\boldsymbol{Y}_t - \boldsymbol{f}_t)\right)$$

where $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$ are obtained with the DLM forward filtering equations and D_t denotes the information available up to time t.

With an inverse gamma prior with parameters a_σ and b_σ for σ^2, $p(\sigma^2|\lambda, D_T)$ is an inverse gamma distribution with shape parameter $T/2 + a_\sigma$ and scale parameter $\sum_{t=1}^{T}(\boldsymbol{Y}_t - \boldsymbol{f}_t)'\boldsymbol{Q}_t^{-1}(\boldsymbol{Y}_t - \boldsymbol{f}_t)/2 + b_\sigma$. Also

With an inverse gamma prior with parameters a_σ and b_σ for σ^2, $p(\sigma^2|\lambda, D_T)$ is an inverse gamma distribution with shape parameter $T/2 + a_\sigma$ and scale parameter $\sum_{t=1}^{T}(\boldsymbol{Y}_t - \boldsymbol{f}_t)'\boldsymbol{Q}_t^{-1}(\boldsymbol{Y}_t - \boldsymbol{f}_t)/2 + b_\sigma$. Also

$$p(\lambda|D_T) \propto p(\lambda) \prod_{t=1}^{T} |\boldsymbol{Q}_t|^{-1/2} \left(\sum_{t=1}^{T} (\boldsymbol{Y}_t - \boldsymbol{f}_t)'\boldsymbol{Q}_t^{-1}(\boldsymbol{Y}_t - \boldsymbol{f}_t) + b_\sigma \right)^{-T/2+a\sigma}$$

To perform the MCMC follow the algorithm:

To perform the MCMC follow the algorithm:

- Run the forward filtering equations to obtain $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$.

To perform the MCMC follow the algorithm:

- Run the forward filtering equations to obtain $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$.
- Choose a jumping distribution for λ, say, $q(\lambda^*|\lambda)$ and sample λ^*.

To perform the MCMC follow the algorithm:

- Run the forward filtering equations to obtain $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$.
- Choose a jumping distribution for λ, say, $q(\lambda^*|\lambda)$ and sample λ^*.
- Let

$$\alpha = \min\left\{1, \frac{p(\lambda^*)q(\lambda|\lambda^*)}{p(\lambda)q(\lambda^*|\lambda)}\right\}$$

To perform the MCMC follow the algorithm:

- Run the forward filtering equations to obtain $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$.
- Choose a jumping distribution for λ, say, $q(\lambda^*|\lambda)$ and sample λ^*.
- Let

$$\alpha = \min\left\{1, \frac{p(\lambda^*)q(\lambda|\lambda^*)}{p(\lambda)q(\lambda^*|\lambda)}\right\}$$

- If λ^* is rejected, then go the next iteration. If it is accepted then sample σ^{2*} from $p(\sigma^2|\lambda^*, D_T)$ and $\boldsymbol{\beta}^*$ using backward sampling.

To perform the MCMC follow the algorithm:

- Run the forward filtering equations to obtain $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$.
- Choose a jumping distribution for λ, say, $q(\lambda^*|\lambda)$ and sample λ^*.
- Let

$$\alpha = \min\left\{1, \frac{p(\lambda^*)q(\lambda|\lambda^*)}{p(\lambda)q(\lambda^*|\lambda)}\right\}$$

- If λ^* is rejected, then go the next iteration. If it is accepted then sample σ^{2*} from $p(\sigma^2|\lambda^*, D_T)$ and $\boldsymbol{\beta}^*$ using backward sampling.

The former avoids sampling large numbers of σ^2 and β that would be rejected anyway.

Temperature Trends

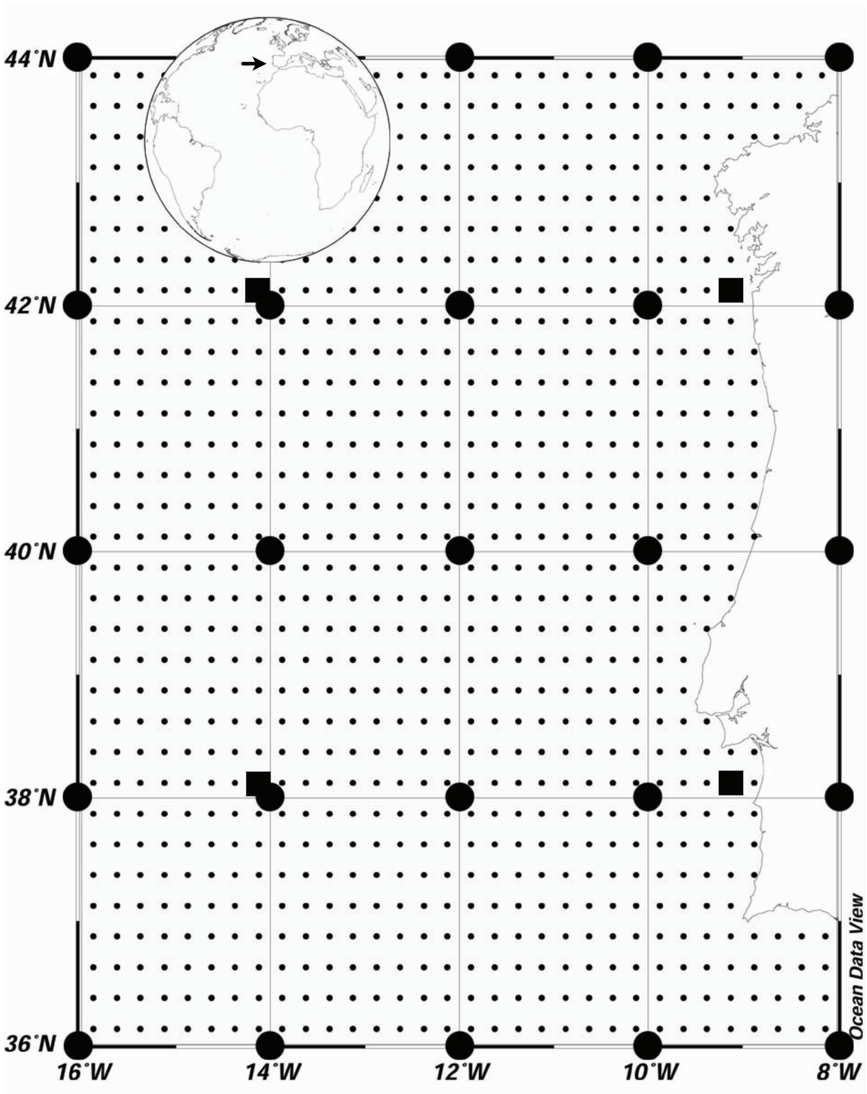
A DCLM spatio-temporal model is developed for the study of trends in the temperature of the Atlantic Ocean off the coast of Portugal.

The goal is to use last century temperature records to describe possible trends. The aim of the model is to quantify the variation of such trends with respect to location and depth.

The region of interest is Atlantic Ocean off the coast of the Iberian Peninsula. The data consist of nearly 260 000 monthly 1/4°grid ocean temperature means at 14 standard depths (0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400 and 500 m), computed from World Ocean Database 2001, using records between Jan-1901 and Dec-2000.

We use the differences between the observations and the monthly 1/4° grid climatological mean fields from the World Ocean Atlas 2001. These are usually referred to as anomalies.

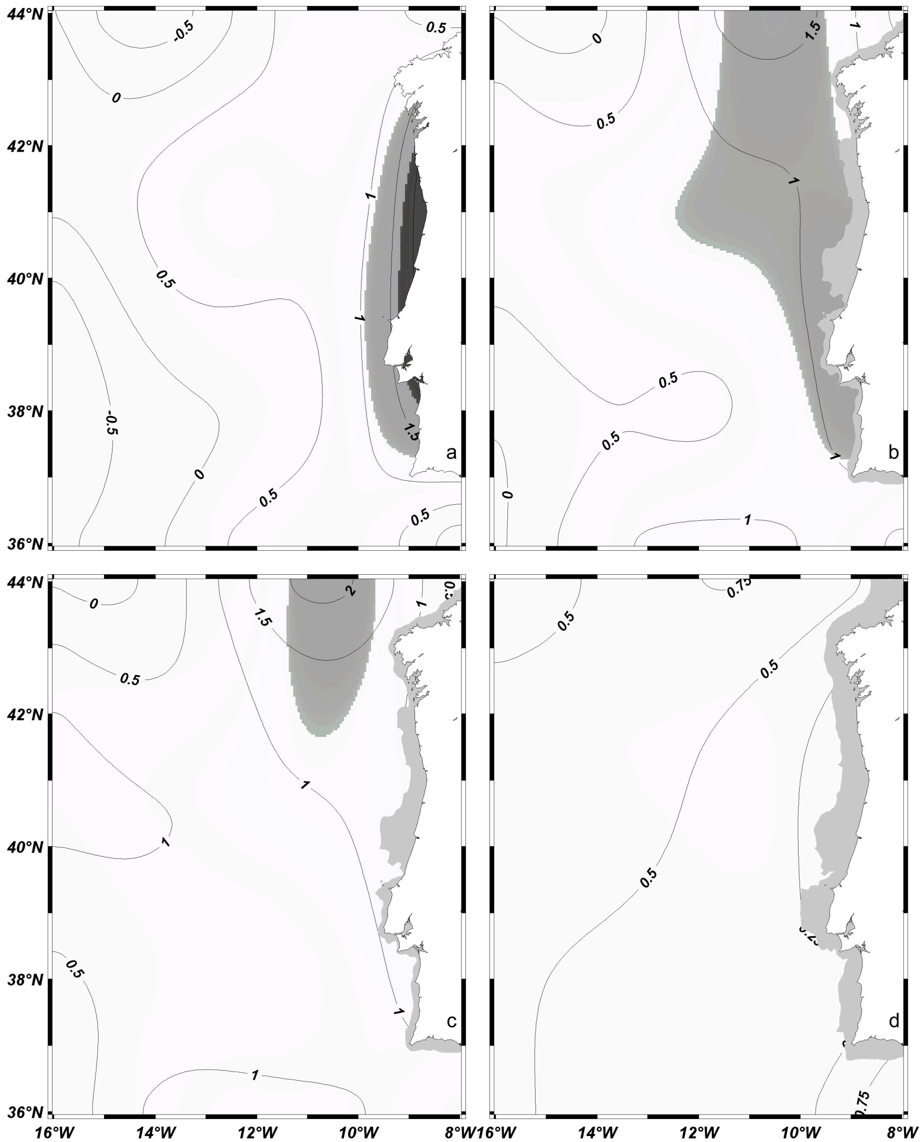
The 1/4° grid used to compute temperature anomalies is represented by small dots. Large dots denote the surface 2° grid lattice for the kernel centers.



Letting $\boldsymbol{Y}_t$ be the vector of n_t observations made in month t, we fit the model

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{X}_t(\alpha_t + \gamma(t - t_R)) + \epsilon_t, & \varepsilon_t &\sim N(0, \sigma^2 I_t), \\ \alpha_t &= \alpha_{t-1} + w_t, & w_t &\sim N(0, W_t). \end{aligned}$$

$\boldsymbol{X}_t$ is a $n_t \times J$ convolution matrix containing the convolution weights. These are obtained from a 3D Gaussian kernel. α_t [°C] and β [°C/month] are $J \times 1$ vectors associated with the baselines and the linear trends. $J = 75$. W_t is defined via discount factors.



Long-term trend estimates [$^\circ$C(100yr)$^{-1}$] at (a) 0 m, (b) 150 m, (c) 200 m, (d) 500 m. Intermediate and dark gray regions have 95% posterior intervals above 0 and 0.5 $^\circ$C(100yr)$^{-1}$, respectively. Light gray regions are shallower than the depth analyzed.

SVD Decomposition

The Singular Value Decomposition, referred to as SVD of a matrix $A \in \mathbb{R}^{m \times n}$, for arbitrary dimensions n and m is given as

$$A = U \Sigma V^T$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times m}$ are orthogonal matrices. $\Sigma \in \mathbb{R}^{m \times n}$ is such that $\sigma_{ij} = 0$ if $i \neq j$ and $\sigma_{ii} = \sigma_i \geq 0$. The non-zero elements of Σ are in decreasing order and are knowns as the **singular values of A**.

The singular values correspond to the positive square roots of the eigenvalues of $A^T A$. The columns of U and V are, respectively, the eigenvectors of AA^T and $A^T A$.

The **condition** of a full rank $m \times m$ matrix A can be defined as

$$\mathrm{cond}(A) = \frac{\sigma_1}{\sigma_m}$$

This number gives an indication of how close to being linearly dependent the columns of the matrix A are. Clearly $\mathrm{cond}(A) \geq 1$. If A is orthogonal, $\mathrm{cond}(A) = 1$. Thus "close to independent columns" correspond to a condition close to 1.

The **pseudoinverse** X of a matrix A satisfies: (1) $AXA = A$; (2) $XAX = X$; (3) AX is symmetric; (4) XA is symmetric. We can obtain a pseudoninverse of a matrix $A = U\Sigma V^T$ as

$$\tilde{A} = V\tilde{\Sigma}U^T$$

where $\tilde{\sigma_i} = 1/\sigma_i$, if $\sigma_i > 0$ and zero otherwise.

Obtaining the EOF Representation

Consider a spatio-temporal process $Y(s,t)$ observed at n locations and T time points, with trends and seasonality removed. We arrange the data in a $n \times T$ dimensional array, say Y. We obtain the SVD decomposition $Y = U\Sigma V^T$. Thus we have that,

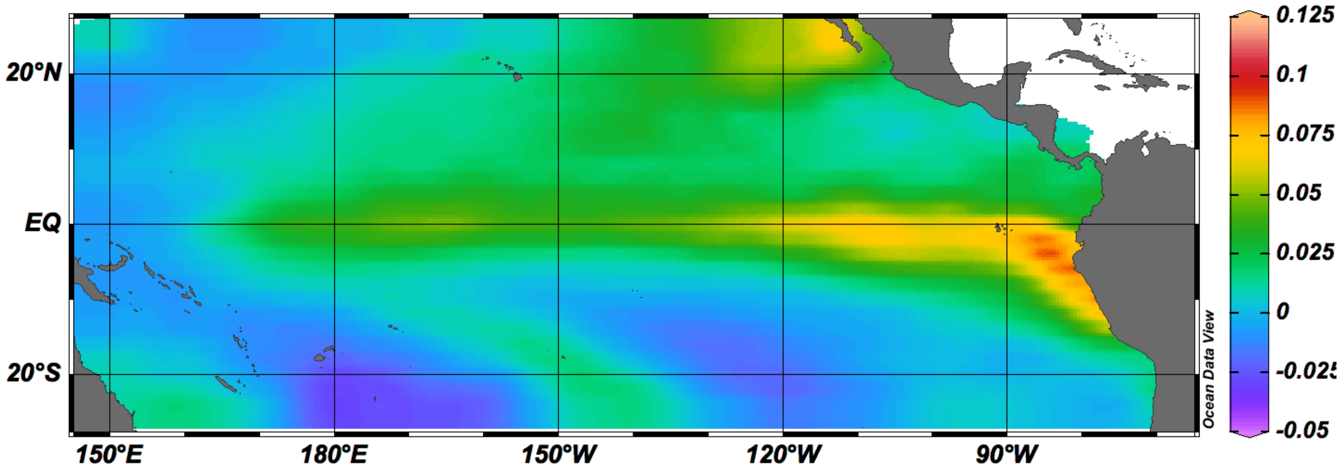
$$Y(s_i, t_j) = \sum_{l=1}^{n \wedge T} u_l(s_i)\sigma_l v_l(t_j)$$

provides a discrete K-L representation of the spatio-temporal field.

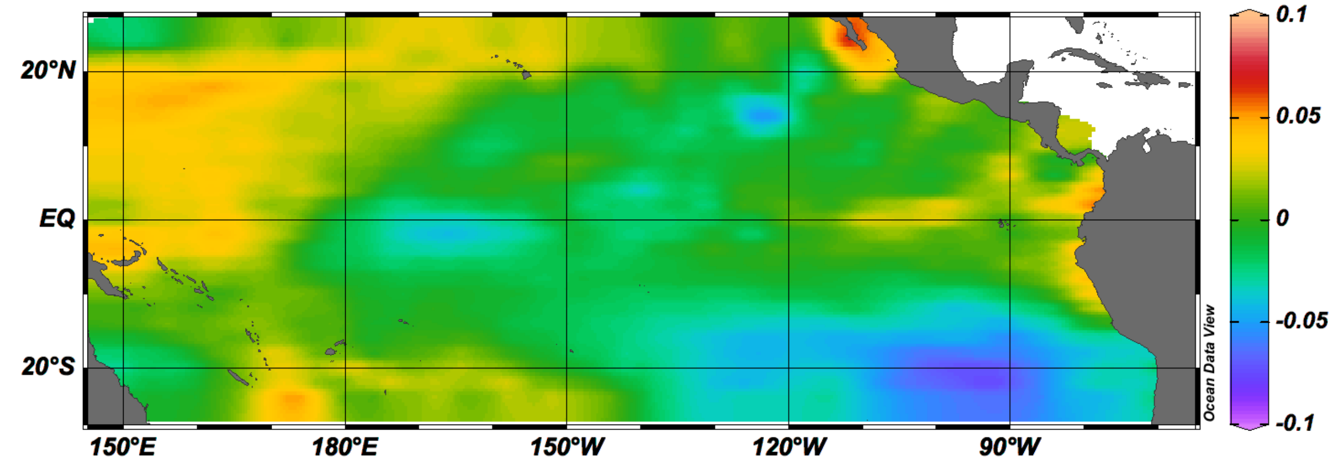
Notice that, if $Y(s,t)$ is stationary in time, $YY^T = U\Sigma^2 U^T$ provides an estimate of the spatial covariance matrix, and the column of U provide estimates for an orthogonal basis of eigenvectors.

First Four EOF of Tropical Pacific Sea Surface Temperature

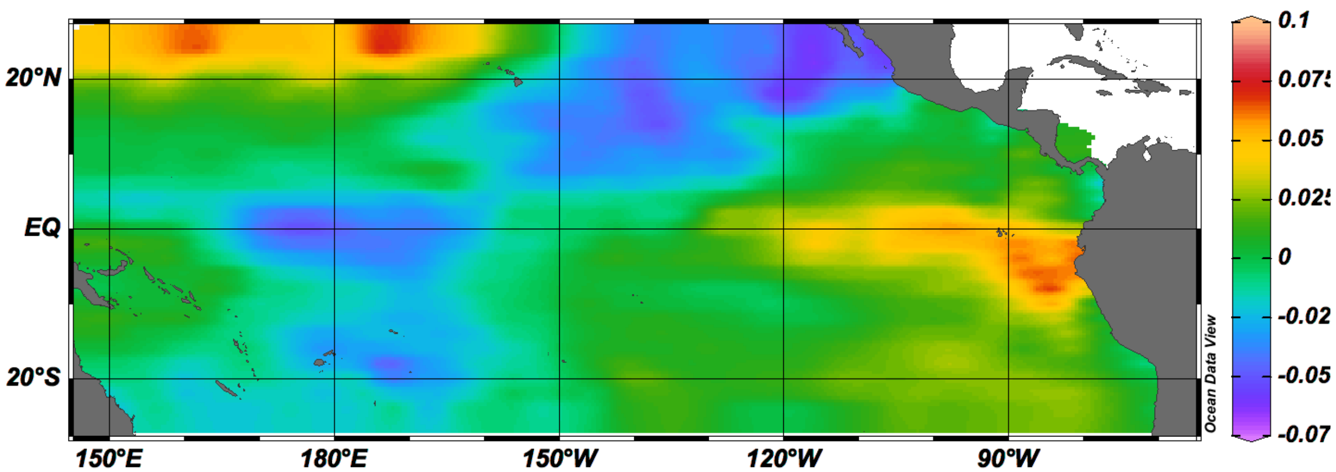
EOF 1



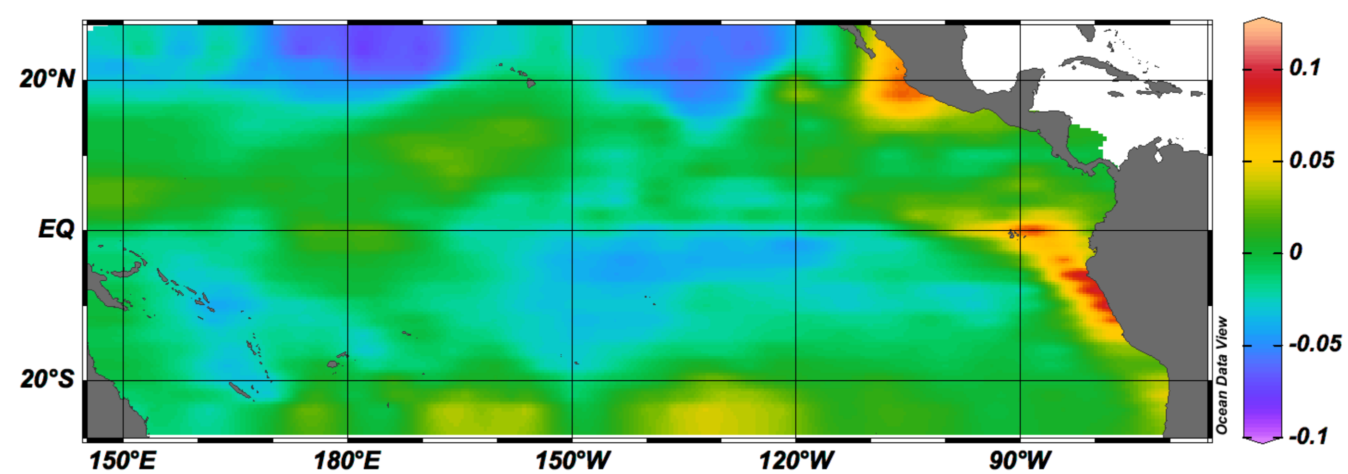
EOF 2



EOF 3



EOF 4



EOFs obtained from monthly SST data from Sept 2013 to Feb 2016

Coefficients of the First Four EOFs