A Bayesian Hierarchical Model

The general structure of a Bayesian hierarchical dynamic model for a vector of time-varying observations irregularly scattered in space is

$$Y_t = H_t X_t + \varepsilon_t, \;\; \varepsilon_t \sim N(0, V_t)$$ Observation Equation

$$X_t = \mathcal{M}(X_{t-1}, \theta) + \omega_t, \;\; \omega_t \sim N(0, W_t)$$ System Equation

A Bayesian Hierarchical Model

The general structure of a Bayesian hierarchical dynamic model for a vector of time-varying observations irregularly scattered in space is

$$Y_t = H_t X_t + \varepsilon_t, \;\; \varepsilon_t \sim N(0, V_t)$$ Observation Equation

$$X_t = \mathcal{M}(X_{t-1}, \theta) + \omega_t, \;\; \omega_t \sim N(0, W_t)$$ System Equation

- The observation equation links the data to the state vector using an ``incidence matrix"
- The evolution allows for non-linear propagation between subsequent times

Linear Dynamic BHM

In the previous BHM the state vectors usually correspond to locations on a regular grid.

- The incidence matrix has 0 or 1 entries that assign the observations a corresponding grid cell.
- Notice that, due to this fact, the physical units of the state vector and the observations are the same.

Linear Dynamic BHM

In the previous BHM the state vectors usually correspond to locations on a regular grid.

The incidence matrix has 0 or 1 entries that assign the observations a corresponding grid cell.

Notice that, due to this fact, the physical units of the state vector and the observations are the same.

$$\boxed{\begin{aligned} Y_t &= H_t X_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, V) \\ X_t &= M X_{t-1} + \omega_t, \quad \omega_t \sim N(0, W) \end{aligned}}$$

$$Y_t \in \mathbb{R}^n, \quad X_t \in \mathbb{R}^p$$
$$V \in \mathbb{R}^{n \times n}, \quad M, W \in \mathbb{R}^{p \times p}$$
$$t = 1, \ldots, T$$

$$\begin{aligned} \mathrm{vec}(M) &\sim N(\mu, \Sigma) \\ V^{-1} &\sim Wish((\nu_V C_V)^{-1}, \nu_V) \\ W^{-1} &\sim Wish((\nu_W C_W)^{-1}, \nu_W), \end{aligned}$$

A linear evolution BHM with completely general evolution matrix

MCMC for the BHM

Estimation for the proposed BMH with linear evolution can be achieved using MCMC that consists only of Gibbs sampling steps.

MCMC for the BHM

Estimation for the proposed BMH with linear evolution can be achieved using MCMC that consists only of Gibbs sampling steps.

Conditional on the matrices V, W and M, the model corresponds to a dynamic linear model. Thus the state vectors can be sampled using forward filtering backwards sampling.

MCMC for the BHM

Estimation for the proposed BMH with linear evolution can be achieved using MCMC that consists only of Gibbs sampling steps.

Conditional on the matrices V, W and M, the model corresponds to a dynamic linear model. Thus the state vectors can be sampled using forward filtering backwards sampling.

To sample V and W we use conjugacy to find that the full conditional distributions of those matrices correspond to inverse Wishart distributions.

$$W^{-1} \sim Wish\left(\left(\sum_{t=1}^{T}(X_t - MX_{t-1})(X_t - MX_{t-1})' + \nu_W C_W\right)^{-1}, \nu_w + T\right)$$

With a similar expression for V. Very often it is assumed that V is a multiple of the identity, though.

MCMC for the BHM

Sampling M can also be achieved from a Gibbs step. To obtain the full conditional write the evolution equation in block form as

$$\mathrm{vec}(X_{1:T}) = (X_{0:T-1} \otimes I_n)\mathrm{vec}(M) + \mathrm{vec}(\omega_{1:T})$$

then use conjugacy to obtain that

$$p(\mathrm{vec}(M)|X, V, W) = N(V_m a_m, V_m)$$

where

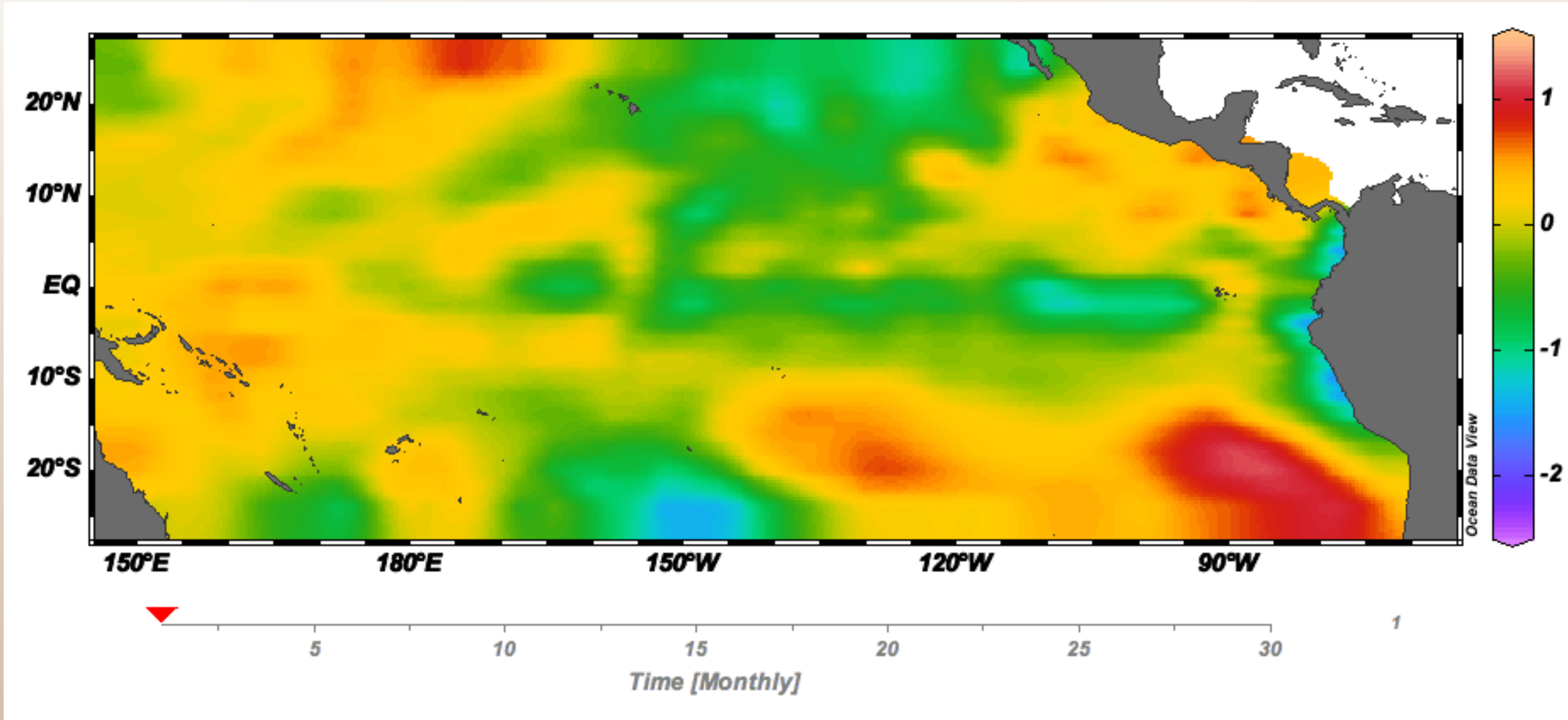
$$V_m = ((X_{0:T-1} \otimes I_n)'(I_T \otimes W)^{-1}(X_{0:T-1} \otimes I_n) + \Sigma^{-1})^{-1}$$

and

$$a_m = (X_{0:T-1} \otimes I_n)'(I_T \otimes W)^{-1}(X_{0:T-1} \otimes I_n) + \Sigma^{-1}\mu$$

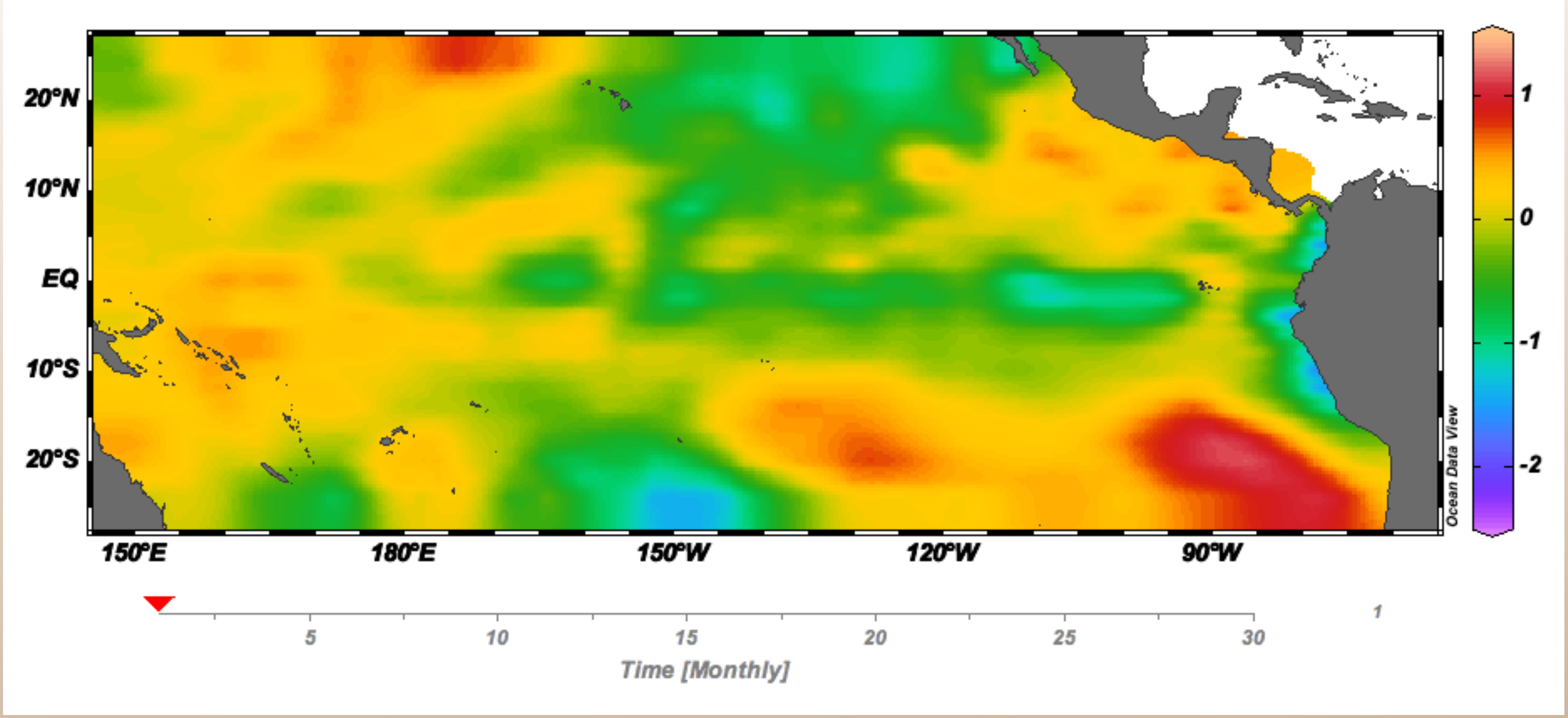
Notice that, for a state vector of a large dimension this is a computationally demanding task.

Sea Surface Temperature



Sea surface temperatures in the tropical Pacific from September 2013 to February 2016. This is the area that is use to study ENSO.

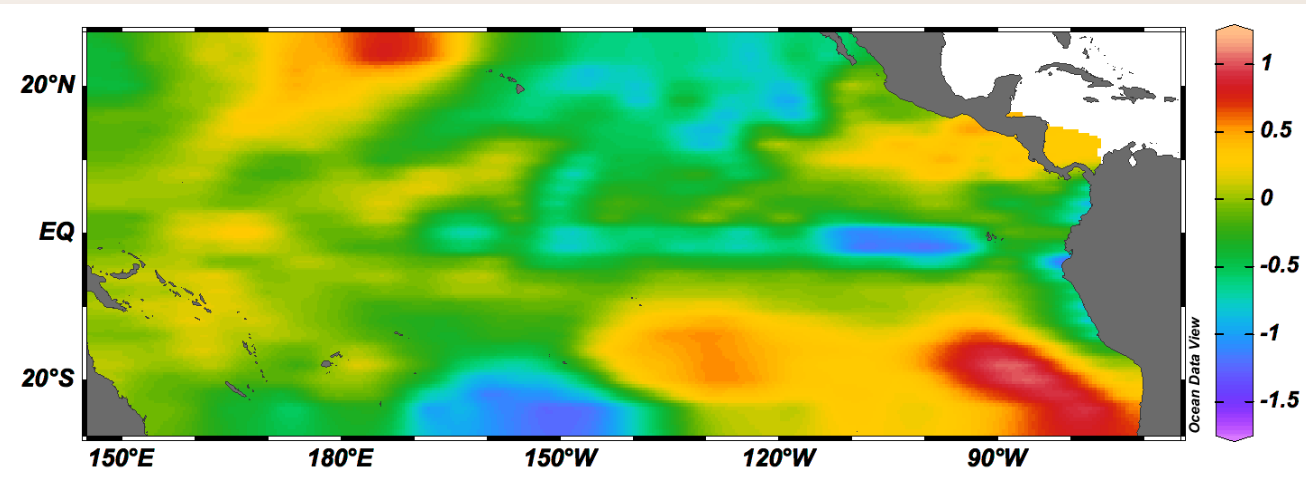
Sea Surface Temperature



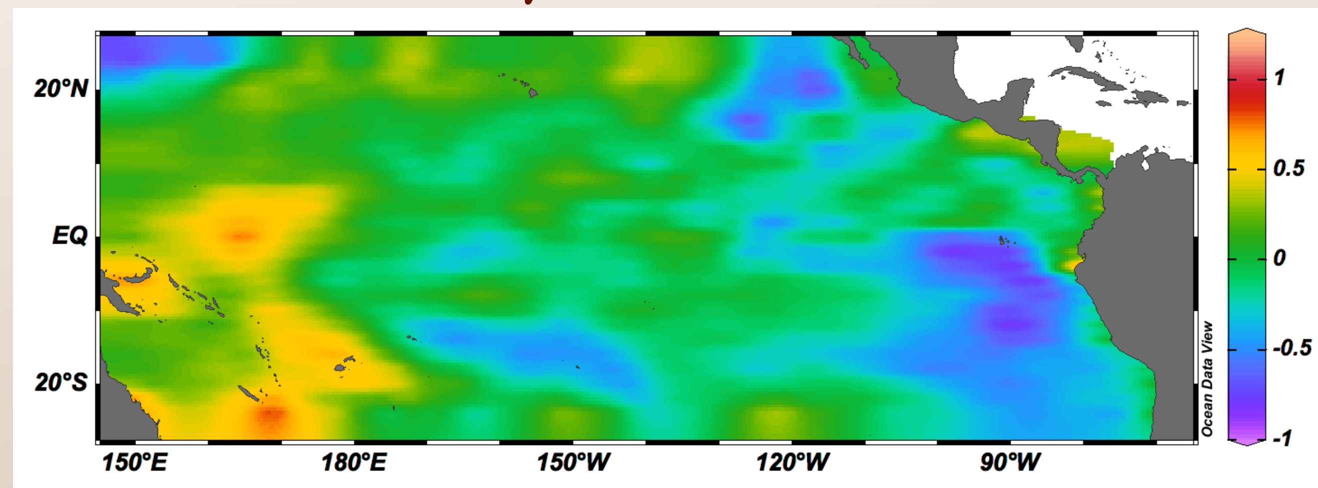
Sea surface temperatures in the tropical Pacific from September 2013 to February 2016. This is the area that is use to study ENSO.

BHM for Tropical Pacific SST

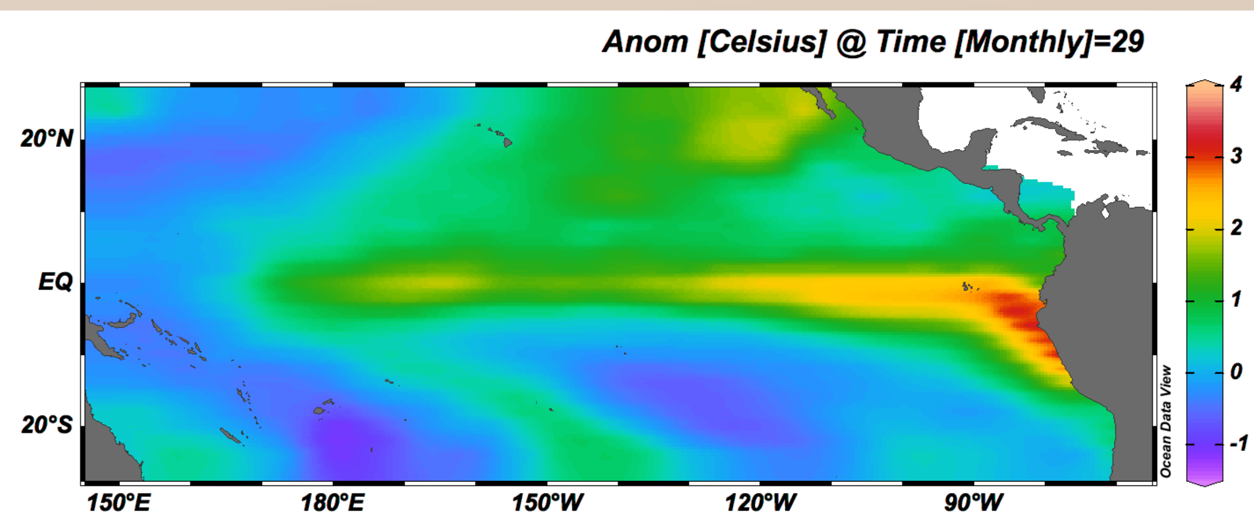
October 2013



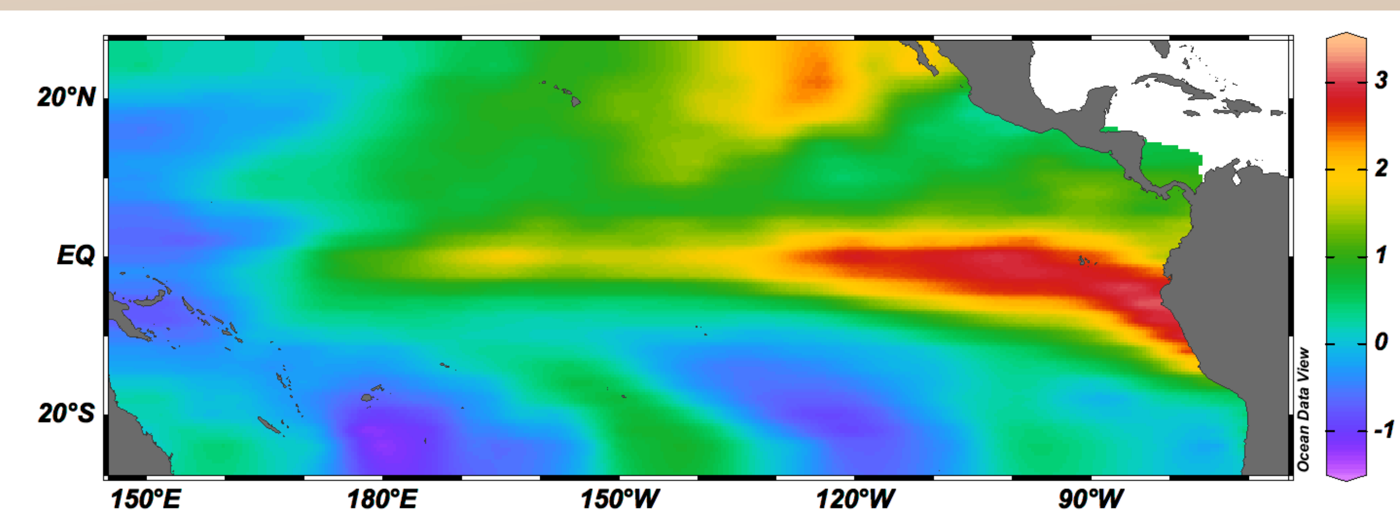
July 2014



February 2016



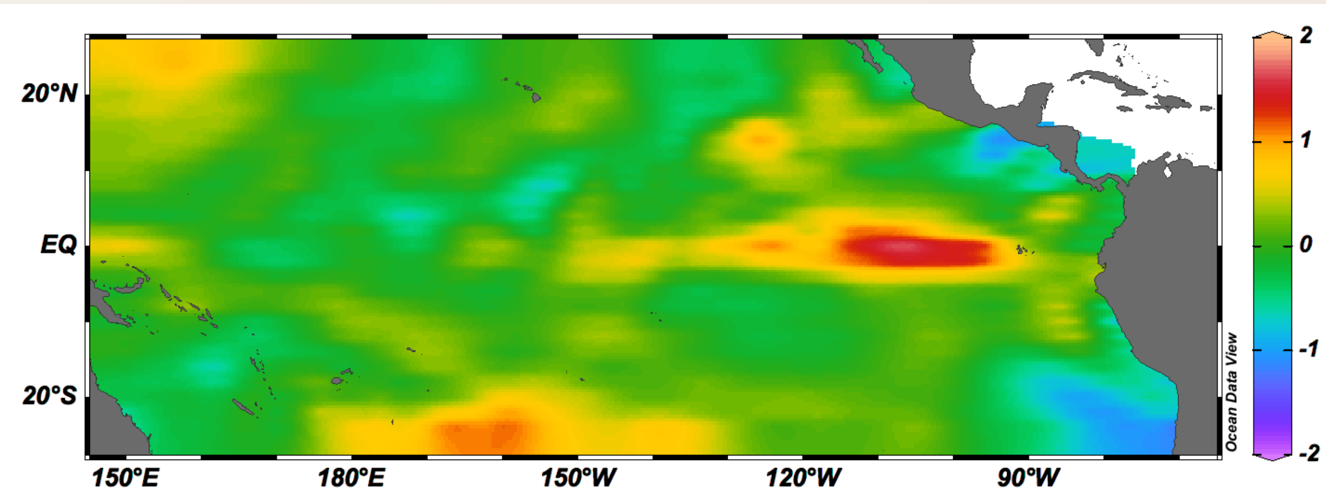
March 2016



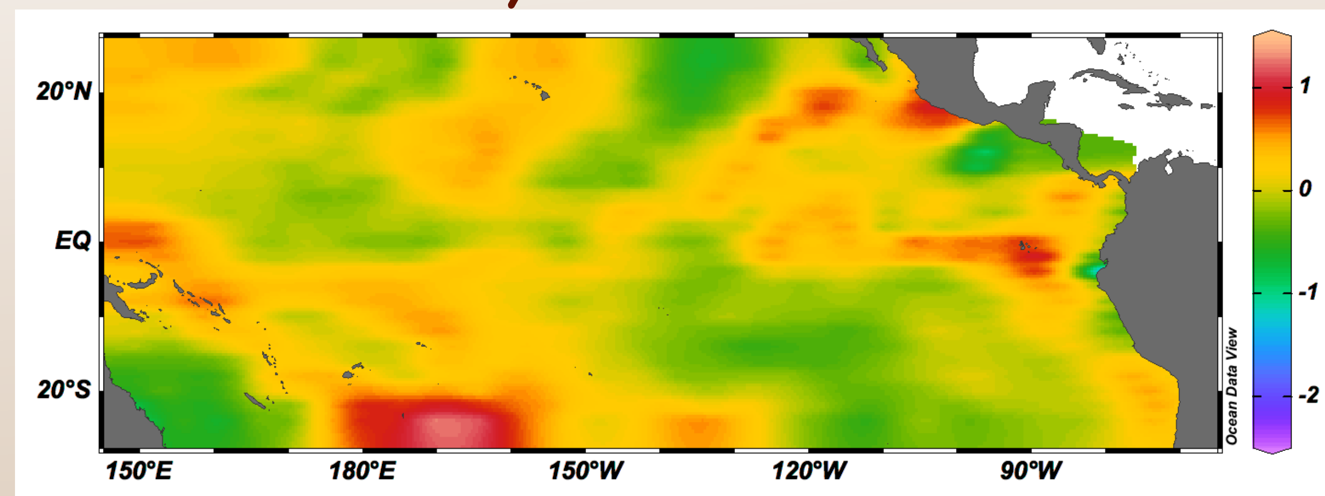
One step ahead predictions for tropical Pacific sea surface temperature obtained using the BHM with 15 EOFs.

BHM for Tropical Pacific SST

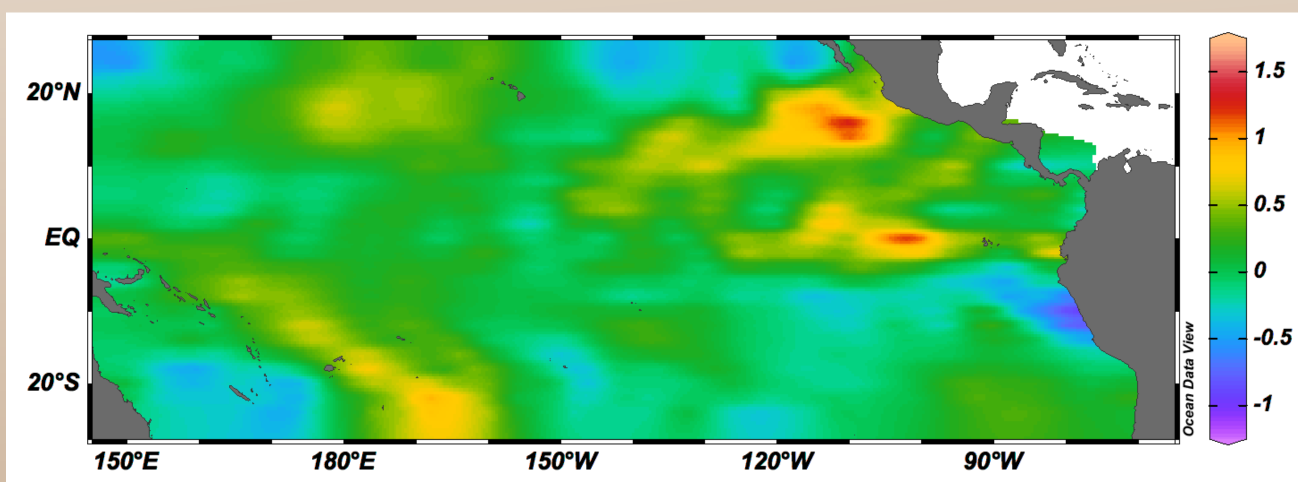
October 2013



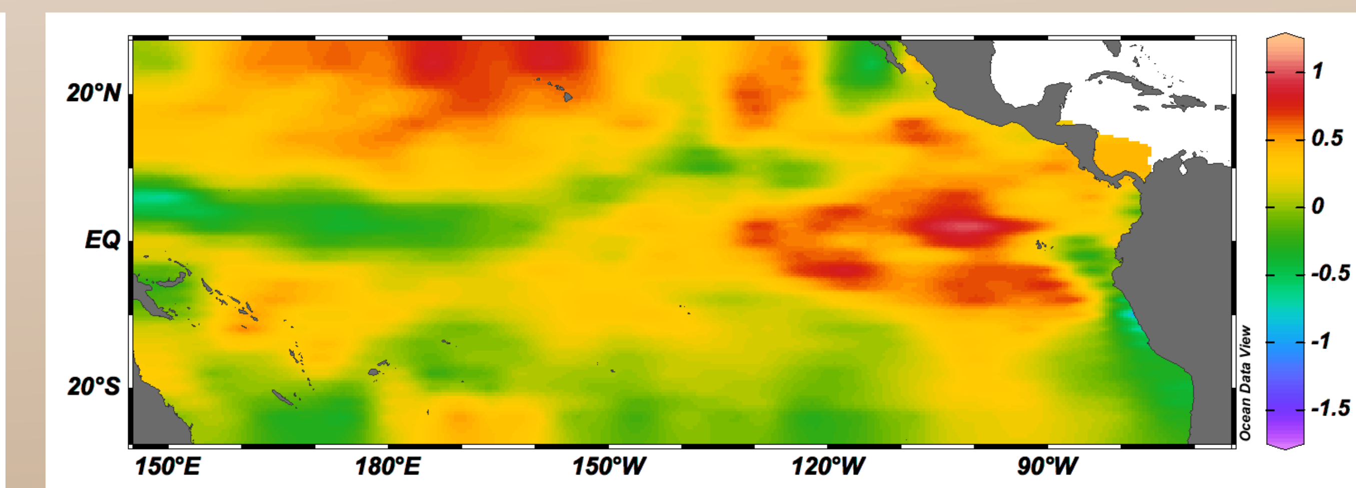
July 2014



May 2015

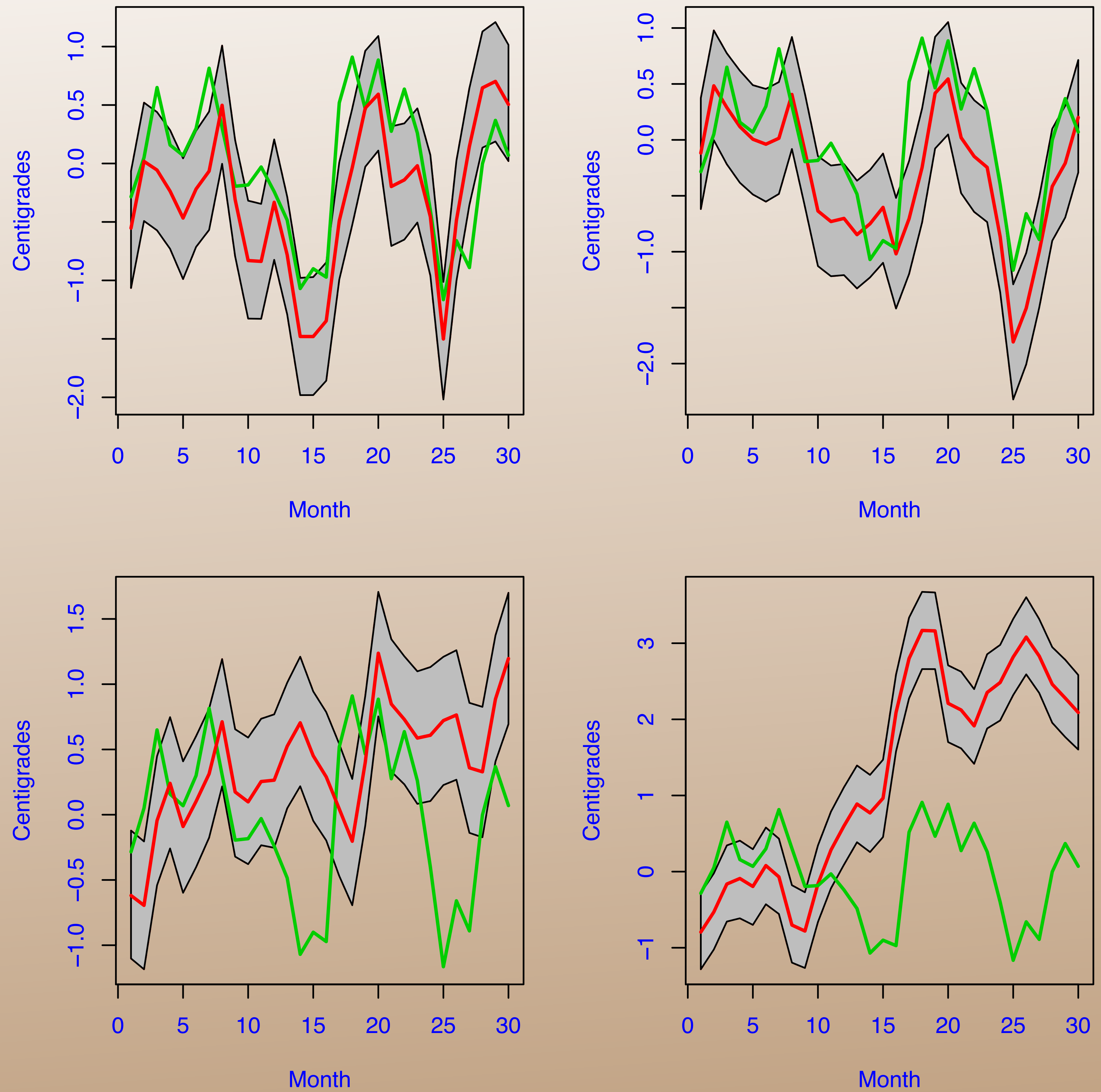


February 2016



One step ahead residuals for tropical Pacific sea surface temperature obtained using the BHM with 15 EOFs.

BHM for Tropical Pacific SST



Time series of one step ahed predictions for four different locations. Grey bands contain 99% probability. Red line is the predictive median. Green line is the actual observation

Integro-Differential Equations

We can obtain a model that is more parsimonious than the general linear dynamics by using a kernel function to describe the spatial redistribution of the process.

An Integro-Differential evolution equation is given as

$$X_t(s) = \int k(u|s, \theta) X_{t-1}(u) du + \omega_t(s)$$

for a family of kernels indexed by a set of parameters θ.

Integro-Differential Equations

We can obtain a model that is more parsimonious than the general linear dynamics by using a kernel function to describe the spatial redistribution of the process.

An Integro-Differential evolution equation is given as

$$X_t(s) = \int k(u|s,\theta)X_{t-1}(u)du + \omega_t(s)$$

for a family of kernels indexed by a set of parameters θ.

The idea of this model is that the process at time t-1 is redistributed in space by the effect of a kernel that may depend on location.

Integro-Differential Equations

We can obtain a model that is more parsimonious than the general linear dynamics by using a kernel function to describe the spatial redistribution of the process.

An Integro-Differential evolution equation is given as

$$X_t(s) = \int k(u|s,\theta) X_{t-1}(u) du + \omega_t(s)$$

for a family of kernels indexed by a set of parameters θ.

The idea of this model is that the process at time t-1 is redistributed in space by the effect of a kernel that may depend on location.

Provided we able to calculate the integral, the inferential problem reduces to the estimation of the kernel parameter.

Properties of IDEs

An IDE with infinitely divisible kernel from a location family satisfies the approximation

$$\frac{\partial X_t(s)}{\partial t} \approx X_t(s) - \mu \frac{\partial X_t(s)}{\partial s} + \frac{1}{2}\sigma^2 \frac{\partial^2 X_t(s)}{\partial s^2} + B_t(s)$$

where the last term corresponds to a Brownian motion and μ and σ are, respectively, the mean and the variance of the kernel.

Properties of IDEs

An IDE with infinitely divisible kernel from a location family satisfies the approximation

$$\frac{\partial X_t(s)}{\partial t} \approx X_t(s) - \mu \frac{\partial X_t(s)}{\partial s} + \frac{1}{2}\sigma^2 \frac{\partial^2 X_t(s)}{\partial s^2} + B_t(s)$$

where the last term corresponds to a Brownian motion and μ and σ are, respectively, the mean and the variance of the kernel.

The effect of an infinite divisible kernel in one unit of time can be decomposed into the sum of n kernels operating in 1/n unit of time.

Properties of IDEs

An IDE with infinitely divisible kernel from a location family satisfies the approximation

$$\frac{\partial X_t(s)}{\partial t} \approx X_t(s) - \mu \frac{\partial X_t(s)}{\partial s} + \frac{1}{2}\sigma^2 \frac{\partial^2 X_t(s)}{\partial s^2} + B_t(s)$$

where the last term corresponds to a Brownian motion and μ and σ are, respectively, the mean and the variance of the kernel.

The effect of an infinite divisible kernel in one unit of time can be decomposed into the sum of n kernels operating in 1/n unit of time.

From the approximation we have that the mean and the variance of the kernel correspond to the advection and the diffusion, respectively of the process. Thus, the effect of the kernel has a physical interpretation.

Calculating the Integral of an IDE

To calculate the redistribution integral for an IDE we consider a set of orthogonal basis functions $\varphi_i(s)$ and use it to represent both the kernel and the process. Let

$$X_t(s) = \sum_{i=1}^{\infty} \varphi_i(s) a_i(t), \quad k(u - s|\theta) = \sum_{j=1}^{\infty} \varphi_j(u) b_j(s, \theta)$$

then, using orthogonality,

$$\int k(u|s, \theta) X_{t-1}(u) du = \sum_{j=1}^{\infty} b_j(s, \theta) a_j(t-1)$$

DLM Representation of an IDE

In practice the basis function representation needs to be truncated to a finite number of terms. The number of basis needed to effectively represent the kernel is related to the its width. We have that

$$X_t = \sum_{j=1}^{L} b_j(s,\theta)a_j(t) = b(s,\theta)'a_t + \omega_t(s)$$

DLM Representation of an IDE

In practice the basis function representation needs to be truncated to a finite number of terms. The number of basis needed to effectively represent the kernel is related to the its width. We have that

$$X_t = \sum_{j=1}^{L} b_j(s,\theta)a_j(t) = b(s,\theta)'a_t + \omega_t(s)$$

Using matrix notation we have an equation for the whole state vector. As the location of the observations may change with time, Φ may depend on time

$$\Phi a_t = B a_{t-1} + \omega_t$$

$$\Downarrow$$

$$a_t = (\Phi'\Phi)^{-1}\Phi B a_{t-1} + (\Phi'\Phi)^{-1}\Phi\omega_t$$

$$\Downarrow$$

$$a_t = M(\theta)a_{t-1} + \eta_t$$

DLM Representation of an IDE

In practice the basis function representation needs to be truncated to a finite number of terms. The number of basis needed to effectively represent the kernel is related to the its width. We have that

$$X_t = \sum_{j=1}^{L} b_j(s,\theta)a_j(t) = b(s,\theta)'a_t + \omega_t(s)$$

Using matrix notation we have an equation for the whole state vector. As the location of the observations may change with time, Φ may depend on time

$$\Phi a_t = B a_{t-1} + \omega_t$$

$$\Downarrow$$

$$a_t = (\Phi'\Phi)^{-1}\Phi B a_{t-1} + (\Phi'\Phi)^{-1}\Phi\omega_t$$

$$\Downarrow$$

$$a_t = M(\theta)a_{t-1} + \eta_t$$

This corresponds to the evolution equation of a conditional DLM

DLM Representation of an IDE

The observations equation corresponding to the basis function representation of the IDE is given as

$$Y_t = \Phi a_t + \varepsilon_t$$

where the error term accounts for observational error as well as truncation error.

DLM Representation of an IDE

The observations equation corresponding to the basis function representation of the IDE is given as

$$Y_t = \Phi a_t + \varepsilon_t$$

where the error term accounts for observational error as well as truncation error.

In a bounded spatial domain, a useful basis representation is given by the Fourier basis. As an example, we have the coefficients, for a one dimensional Gaussian kernel on $[r_1, r_2]$.

$$b_{2j-1}(s, \mu, \sigma) = \frac{\exp(-\frac{\rho_j^2 \sigma^2}{2}) \cos(\rho_j (s + \mu))}{\sqrt{r}}$$

$$b_{2j}(s, \mu, \sigma) = \frac{\exp(-\frac{\rho_j^2 \sigma^2}{2}) \sin(\rho_j (s + \mu))}{\sqrt{r}}$$

$$r = r_2 - r_1, \quad \rho_j = \frac{2\pi}{r}$$

SST Revisited

Need to run Robert's programs for the SST data using the IDE model

Additional Components

$$\boxed{\begin{aligned} Y_t &= \Phi_t a_t + Z_{1t} + \varepsilon_t \\ &\underbrace{\begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix} = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} Z_{1t-1} \\ Z_{2t-1} \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}}_{Z_t = G Z_{t-1} + v_t} \\ a_t &= M_t(\theta) a_{t-1} + \eta_t \end{aligned}}$$

$$\Downarrow$$

$$\boxed{\begin{aligned} Y_t &= H_t \alpha_t + \varepsilon_t \\ \alpha_t &= N_t(\theta,\omega)\alpha_{t-1} + \eta_t \end{aligned}}$$

where

$$H_t = (\Phi_t, 1, 0)$$
$$\alpha_t = (a_t, Z_t)'$$
$$N_t(\theta,\omega) = \begin{pmatrix} M_t(\theta) & 0 \\ 0 & G(\omega) \end{pmatrix}$$

Given the conditional dynamic linear model representation of an IDE model, we can use the flexibility of DLMs to add components that capture trends and cycles. For example, consider adding a component with a cycle of frequency ω.

Higher Order Moments

For and IDE with an infinite divisible location kernel, the former result is generalized to higher order moments

$$\frac{\partial X_t(s)}{\partial t} \approx \sum_{j=1}^{j} (-1)^j \frac{1}{j!} \kappa_j \frac{\partial^j X_t(s)}{\partial s^j} + B_t(s)$$

where $\varkappa_j$ are cumulants of the kernels.

Higher Order Moments

For and IDE with an infinite divisible location kernel, the former result is generalized to higher order moments

$$\frac{\partial X_t(s)}{\partial t} \approx \sum_{j=1}^{j} (-1)^j \frac{1}{j!} \kappa_j \frac{\partial^j X_t(s)}{\partial s^j} + B_t(s)$$

where $\varkappa_j$ are cumulants of the kernels.

The former result indicates that higher moments of the kernel control additional aspects of the resulting IDE process. For example, the third cumulant control the dispersion.

Higher Order Moments

For and IDE with an infinite divisible location kernel, the former result is generalized to higher order moments

$$\frac{\partial X_t(s)}{\partial t} \approx \sum_{j=1}^{j} (-1)^j \frac{1}{j!} \kappa_j \frac{\partial^j X_t(s)}{\partial s^j} + B_t(s)$$

where $\varkappa_j$ are cumulants of the kernels.

The former result indicates that higher moments of the kernel control additional aspects of the resulting IDE process. For example, the third cumulant control the dispersion.

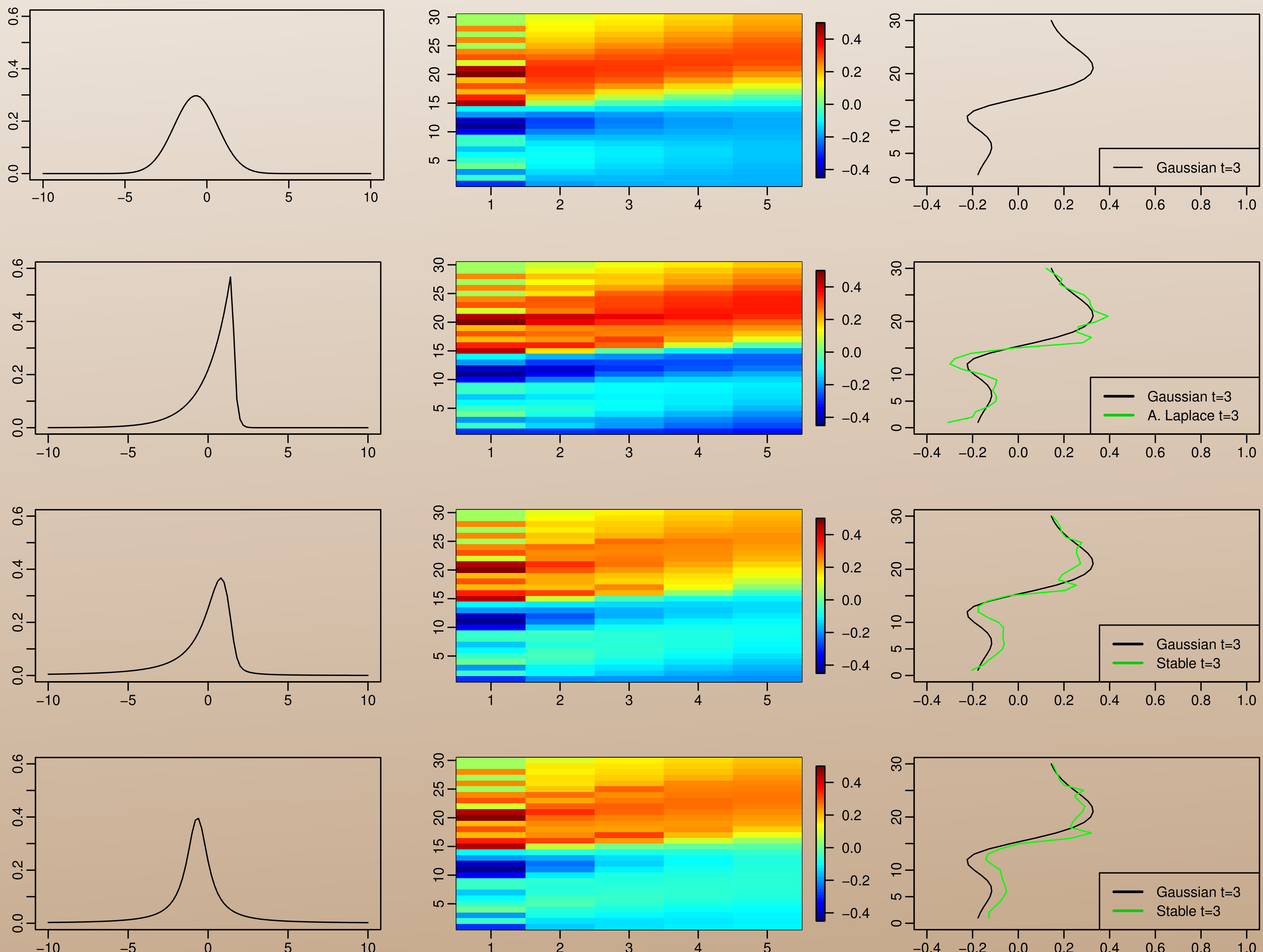
It is possible to obtain an alternative characterization of an IDE process in terms of the hazard function associated with the kernel. This indicates that the kernel tails play a role in the dynamics of the resulting process.

Beyond a Gaussian Kernel

The previous discussion motivates the idea of exploring kernels other than the Gaussian, to add flexibility to the number of different moments and the tail behavior.

Beyond a Gaussian Kernel

The previous discussion motivates the idea of exploring kernels other than the Gaussian, to add flexibility to the number of different moments and the tail behavior.



Simulations of a one dimensional IDE process with different types of kernels (top to bottom): Gaussian, Asymmetric Laplace, Stable with skewness and Stable with no skewness

DLM Representation of an IDE

The fact that it is possible to represent an IDE based spatio-temporal model using the Fourier coefficients of the kernel is a very important property that allows for the use of kernels that are known only through their characteristic function.

DLM Representation of an IDE

The fact that it is possible to represent an IDE based spatio-temporal model using the Fourier coefficients of the kernel is a very important property that allows for the use of kernels that are known only through their characteristic function.

This is the case for kernels in the stable family, that has thick tails and a wide range of skewness, but is defined using the characteristic function:

$$\psi(t) = \exp\{it\mu - |ct|^{\alpha}(1 - i\beta \operatorname{sgn}(t) \tan(\pi\alpha/2))\}$$

This is a four parameter infinitely divisible distribution that includes as special cases the Gaussian and the Cauchy distributions.

Space-Varying IDE Kernels

A space-varying non-stationary behavior for the IDE kernels can be obtained by making the parameters of the kernel varying with space. This entails that the physical properties of the S-T field vary with location.

Space-Varying IDE Kernels

A space-varying non-stationary behavior for the IDE kernels can be obtained by making the parameters of the kernel varying with space. This entails that the physical properties of the S-T field vary with location.

Consider, for example, a field with space-varying advection, due, for example, to the effect of prevailing winds, say Z. We can model μ(s) as a Gaussian random field.

$$\mu(s) = Z_t(s)'\beta_t + \sum_{j=1}^{M} h(s - s_j)\gamma_j$$

$Z_t(s)$ covariates

$h(\cdot)$ process convolution kernel

$(\gamma_1, \ldots, \gamma_M)' \sim N(0, \Sigma)$

Non-Parametric Kernel Representation

To achieve further flexibility in the IDE model we represent the kernel as mixture of Gaussians with fixed variance and location varying mean. The mixing distribution is given a spatial Dirichlet process prior (SDP).

Non-Parametric Kernel Representation

To achieve further flexibility in the IDE model we represent the kernel as mixture of Gaussians with fixed variance and location varying mean. The mixing distribution is given a spatial Dirichlet process prior (SDP).

Truncating the stick-breaking representation of the SDP we obtain the kernel as a finite sum.

$$k(u|s, G_s, \sigma) = \sum_{l=1}^{L} \omega_j \phi(u|s + \mu_l(s), \sigma^2)$$

The weights are obtained from latent beta variables. The vector of atoms corresponds to the Gaussian process that is the base of the random measure.

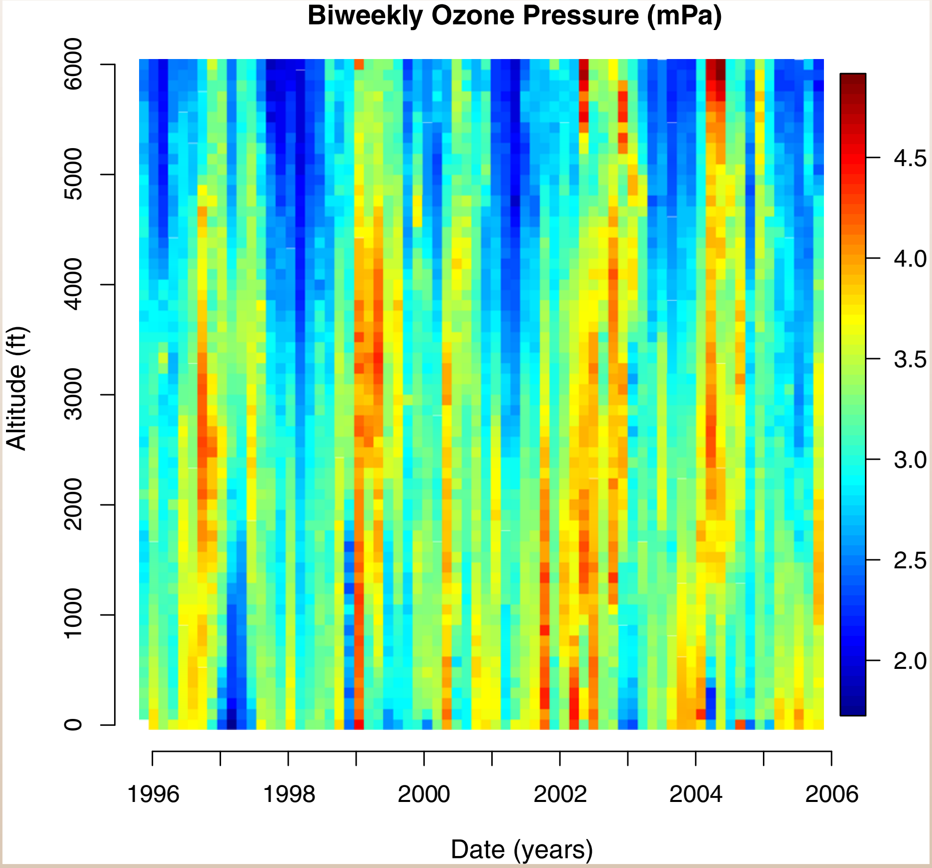
$$\mu_l(s_1), \ldots, \mu_l(s_n)$$

The resulting non-stationary IDE is

$$X_t(s) = \sum_{l=1}^{L} \omega_j \int \phi(u|s + \mu_l(s), \sigma^2) X_{t-1}(u) du + \omega_t(s)$$

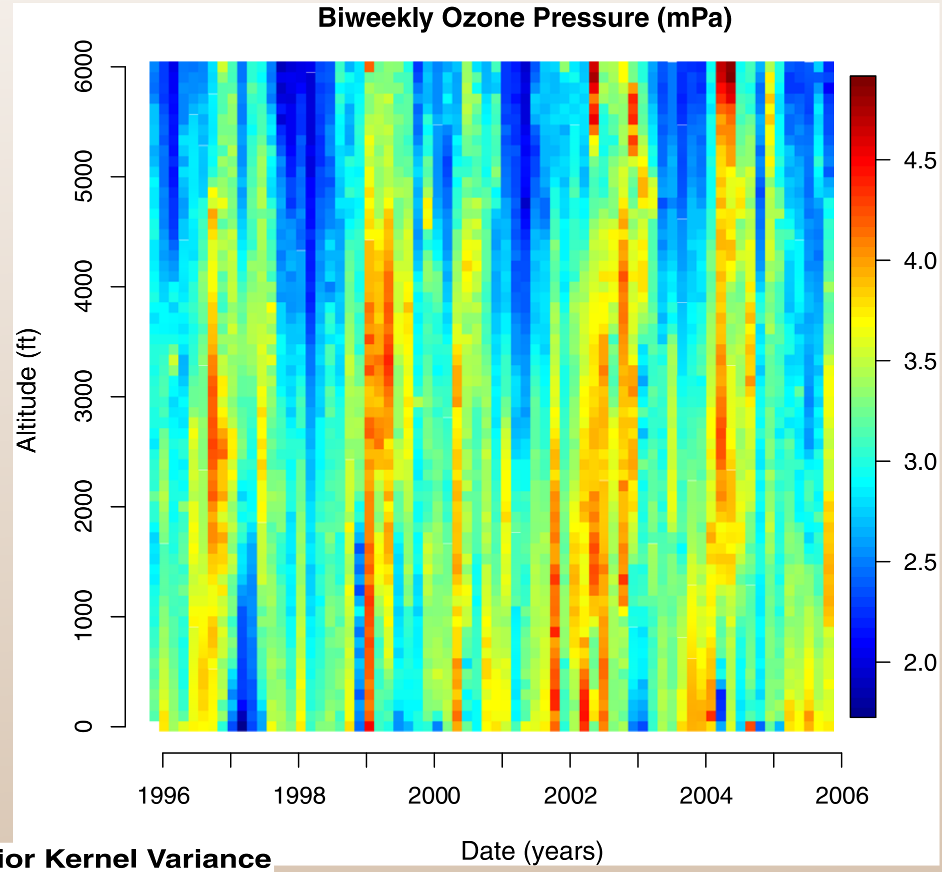
Ozone Data Analysis

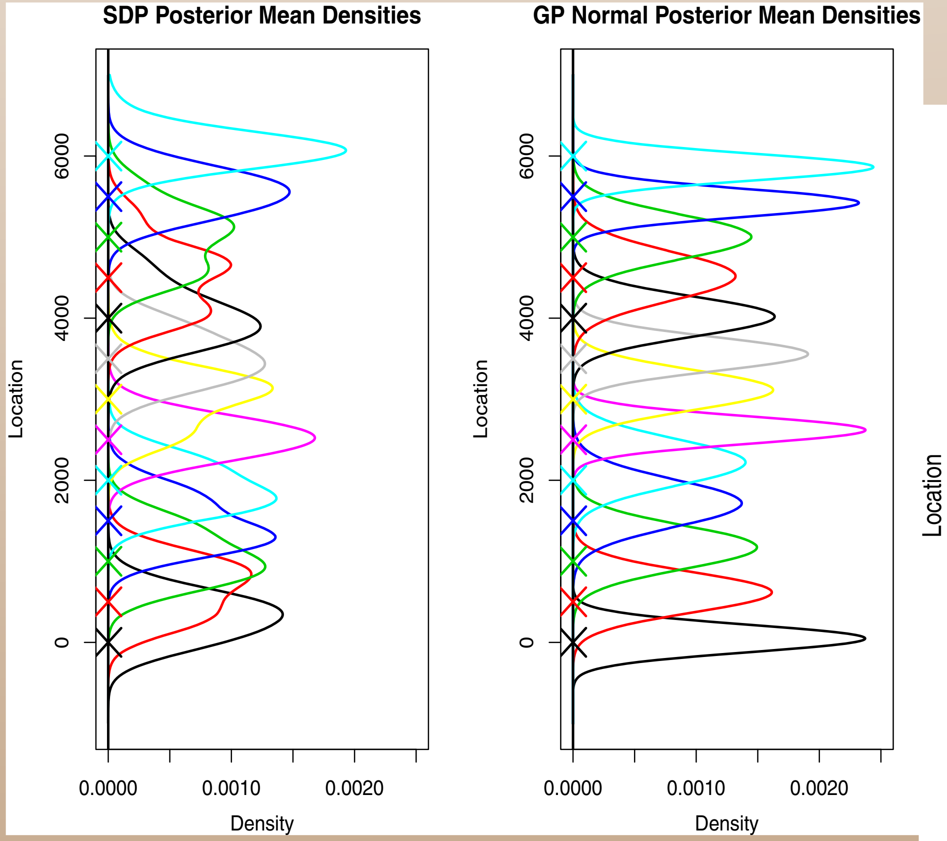
We consider ozone pressure measured by balloons near the North Pole at increasing altitudes. The data were fitted with an IDE plus two time-varying seasonal components.

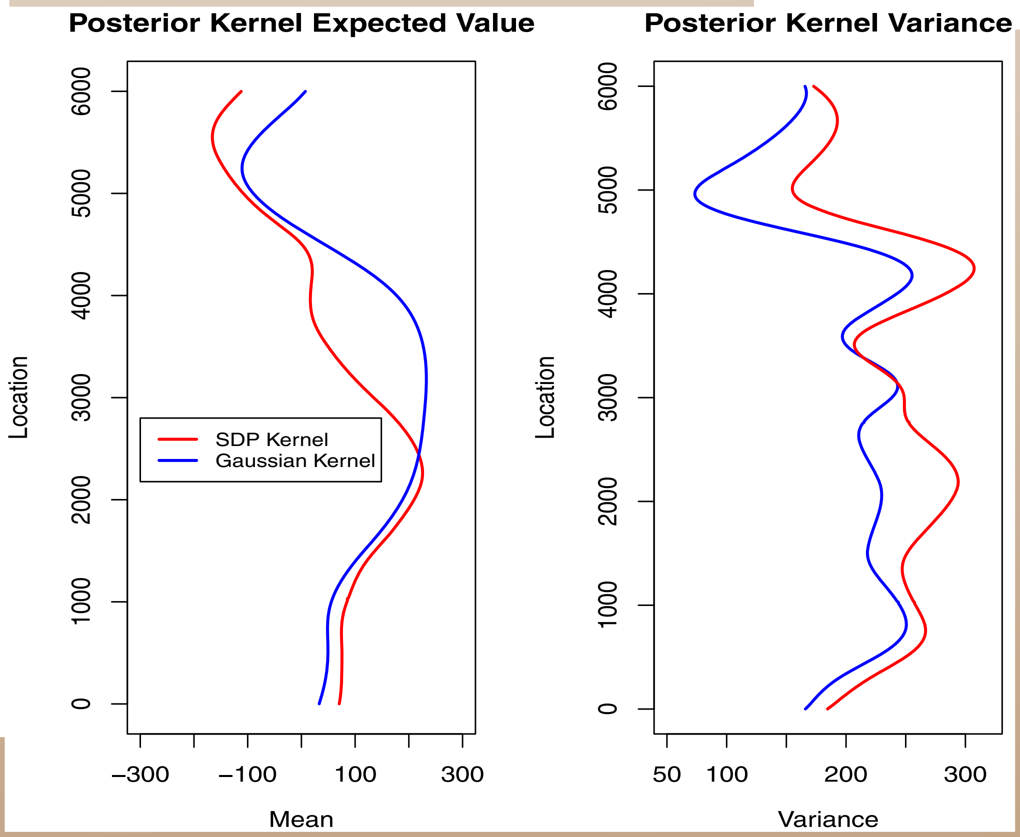


Ozone Data Analysis

We consider ozone pressure measured by balloons near the North Pole at increasing altitudes. The data were fitted with an IDE plus two time-varying seasonal components.



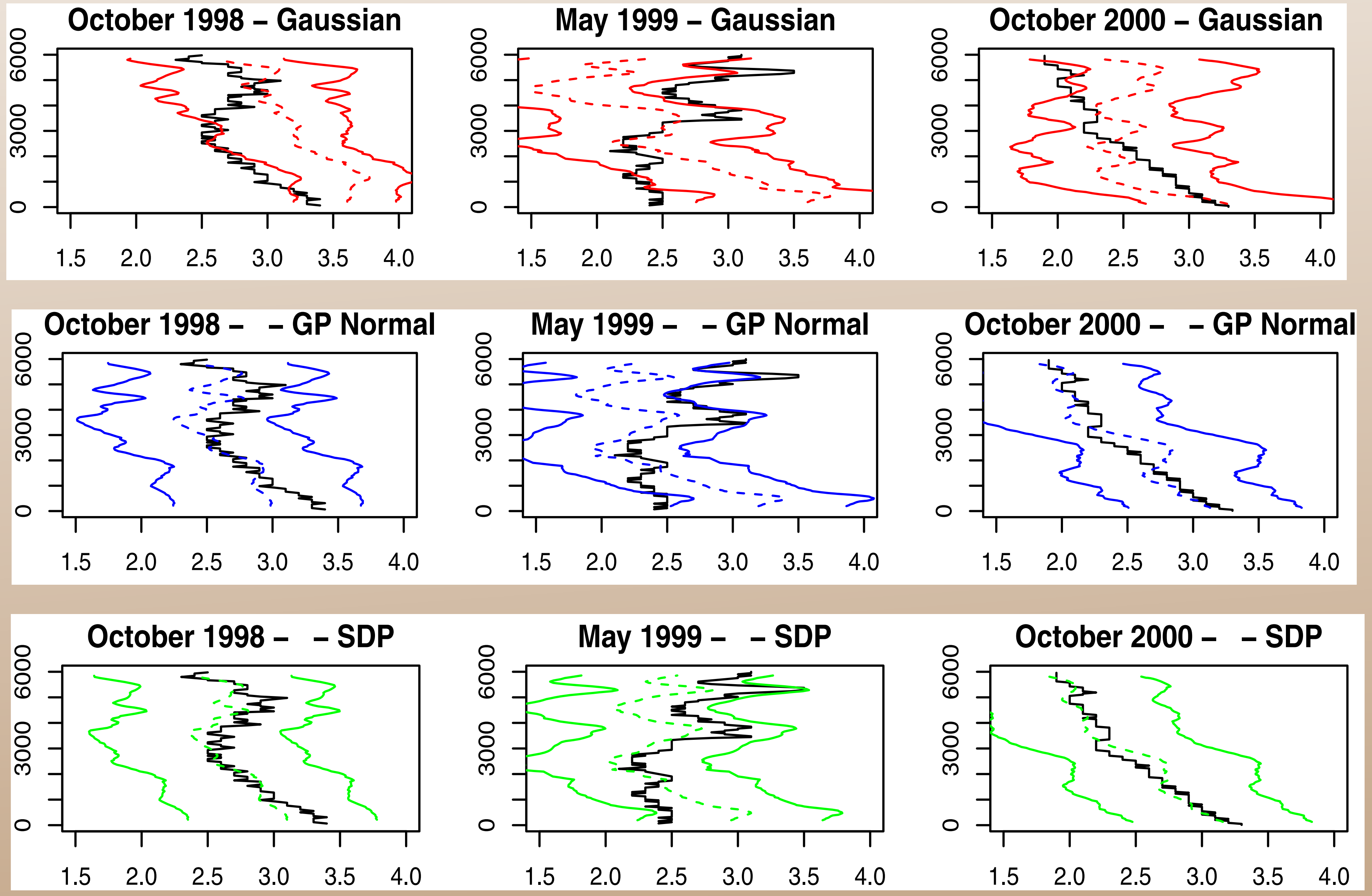




We fitted a Gaussian kernel with space varying mean and variance, and a SDP-based non-parametric kernel

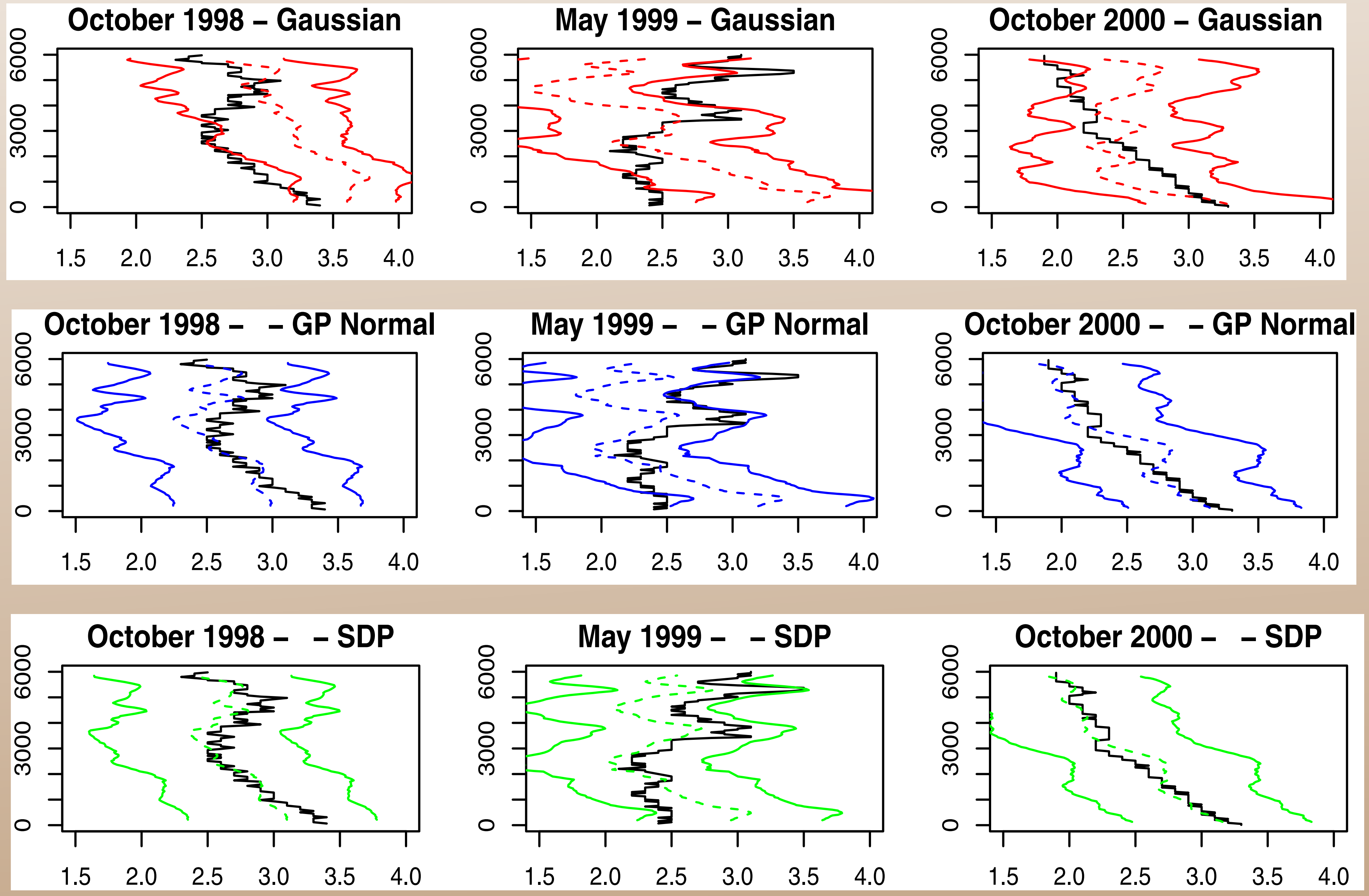
Ozone Data Analysis

The figures show one step ahead predictions for three models: Gaussian kernel with mixed parameters; Gaussian kernels with spatially-varying parameters; SDP-base non-parametric kernel



Ozone Data Analysis

The figures show one step ahead predictions for three models: Gaussian kernel with mixed parameters; Gaussian kernels with spatially-varying parameters; SDP-base non-parametric kernel



The added flexibility of space-varying kernels plays an important role.

Adding kernel shape flexibility improves predictive skill

Extensions to Two Dimensional Space

Extending the results for kernels in the stable family to a two dimensional setting presents the challenge of dealing with the non-parametric representation of the characteristic function

$$\psi(t) = \exp\left\{it'\mu - \int |t'v|^{\alpha}(1 - i\text{sign}(t'v)\tan(\pi\alpha/2))d\Gamma(v)\right\}$$

that depends on the distribution $\gamma = d\Gamma$.

Extensions to Two Dimensional Space

Extending the results for kernels in the stable family to a two dimensional setting presents the challenge of dealing with the non-parametric representation of the characteristic function

$$\psi(t) = \exp\left\{ it'\mu - \int |t'v|^{\alpha}(1 - i\text{sign}(t'v)\tan(\pi\alpha/2))d\Gamma(v) \right\}$$

that depends on the distribution γ = dΓ.

A flexible model for γ is given by a weighted combination of beta densities

$$\gamma(z) = \frac{c}{2\pi}\sum_{m=1}^{M} w_m \text{be}(2/2\pi|m, M, m+1), \quad z \in [0, 2\pi]$$

Spectral Density Model

The weights that define the spectral density are given as

$$w_m = F(2\pi m/M) - F(2\pi(m-1)/M), \;\; m = 1, \ldots, M$$

Where the underlying distribution F is obtained as

$$F = \sum_{j=1}^{\infty} q(1-q)^{j-1}\delta_{x_j}, \quad x_j \sim U[0, 2\pi]$$

Spectral Density Model

The weights that define the spectral density are given as

$$w_m = F(2\pi m/M) - F(2\pi(m-1)/M), \;\; m = 1, \ldots, M$$

Where the underlying distribution F is obtained as

$$F = \sum_{j=1}^{\infty} q(1-q)^{j-1}\delta_{x_j}, \quad x_j \sim U[0, 2\pi]$$

A spatially-varying family of kernels is obtained by assuming that the parameters that define the kernel, μ_1 , μ_2 , log(c) and a legit transformation of q depend on location s.

Spectral Density Model

The weights that define the spectral density are given as

$$w_m = F(2\pi m/M) - F(2\pi(m-1)/M), \;\; m = 1, \ldots, M$$

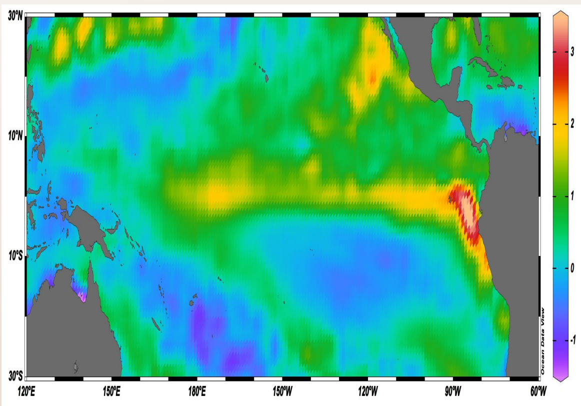
Where the underlying distribution F is obtained as

$$F = \sum_{j=1}^{\infty} q(1-q)^{j-1} \delta_{x_j}, \quad x_j \sim U[0, 2\pi]$$

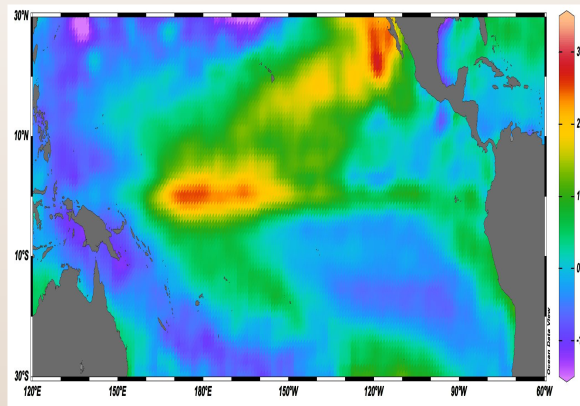
A spatially-varying family of kernels is obtained by assuming that the parameters that define the kernel, μ_1 , μ_2 , log(c) and a legit transformation of q depend on location s.

The spatial dependence of such parameters is modeled using process convolutions. The parameter α is kept fixed for all locations.

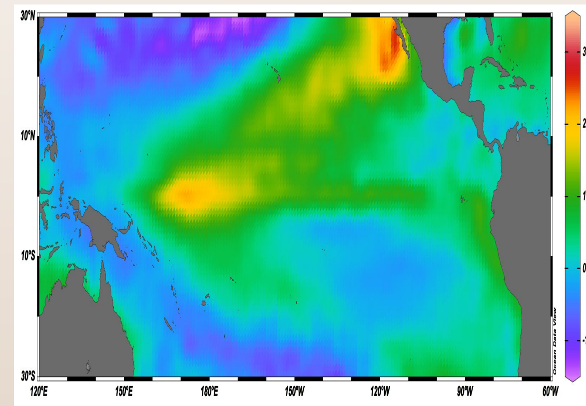
SST Predictions



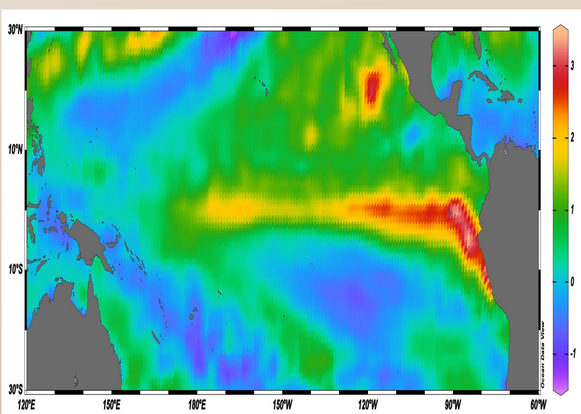
(g) May 2015 Truth



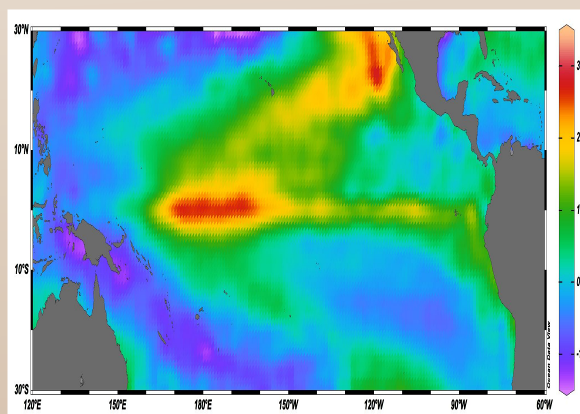
(h) May 2015 Stable Predictions



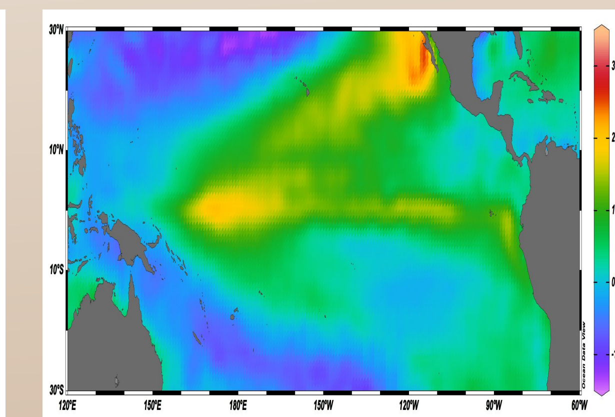
(i) May 2015 Normal Predictions



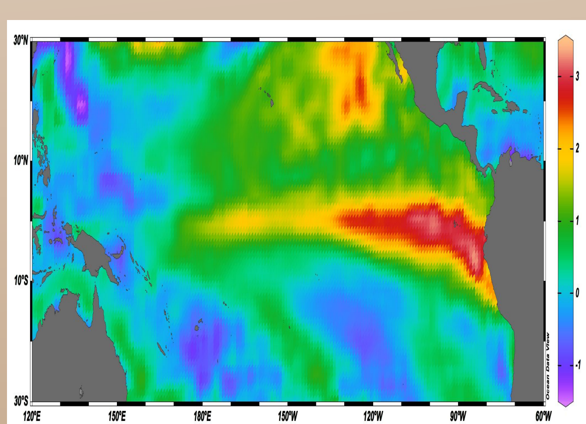
(j) June 2015 Truth



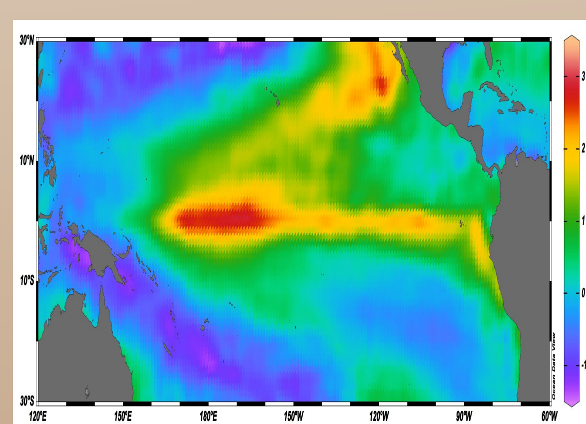
(k) June 2015 Stable Predictions



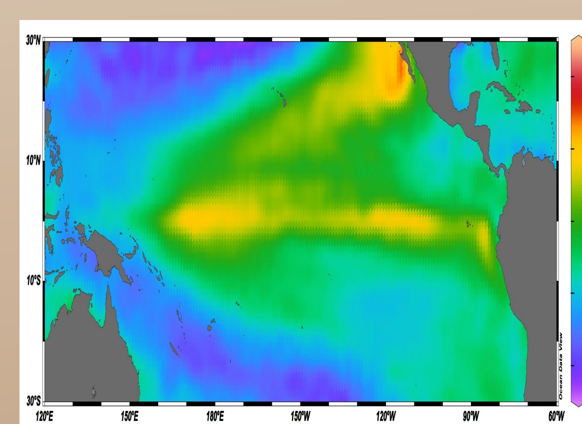
(l) June 2015 Normal Predictions



(m) July 2015 Truth



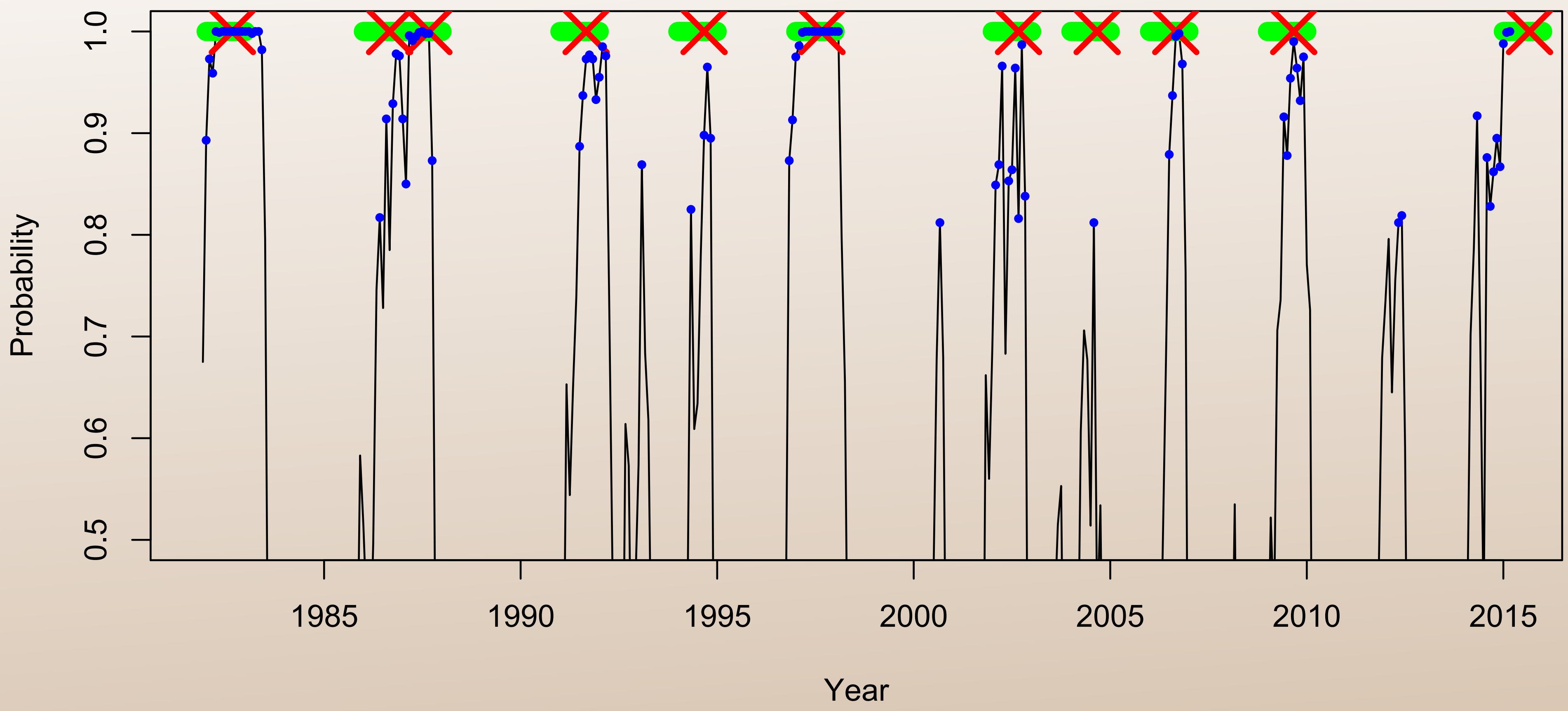
(n) July 2015 Stable Predictions



(o) July 2015 Normal Predictions

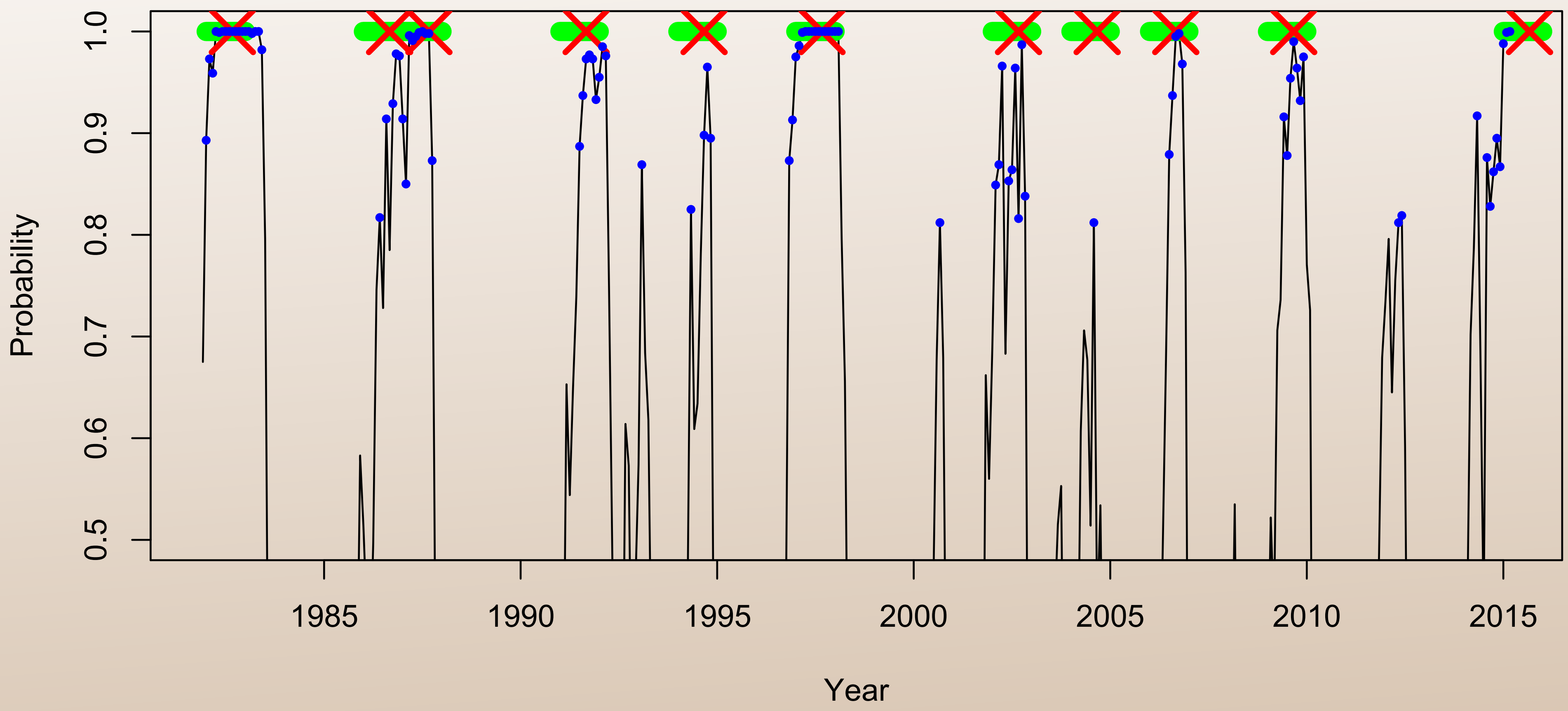
Monthly SST from 12/80 to 3/15 in the El Niño region of the Pacific where used to fit normal and stable spatially-varying IDEs. The predictions for May, June and July are compared to the actual data. These were not used in the fit.

El Niño Predictions



The criterion to declare an El Niño consists of observing an anomaly of more than 3 degrees for more than five months over the region of 120-170 Lon and -5-5 Lat.

El Niño Predictions



We use the IDE model to calculate the probability that an El Niño will occur in the next nine months for each month since 12/1980. The red crosses within green areas correspond to actual El Niño events. Blue dots mark months with probability 0.8 or higher

The criterion to declare an El Niño consists of observing an anomaly of more than 3 degrees for more than five months over the region of 120-170 Lon and -5-5 Lat.

IDE References

- Robert Richardson, Athanasios Kottas and Bruno Sansó (2017) Spatio-Temporal Modelling Using Integro-Difference Equations With Bivariate Stable Kernels

- Robert Richardson, Athanasios Kottas and Bruno Sansó (2017) "Flexible Integro-Difference Equation Modeling for Spatio-Temporal Data". Computational Statistics and Data Anal- ysis, vol. 109, 182–198, http://dx.doi.org/10.1016/j.csda.2016.11.011.

- Robert Richardson, Athanasios Kottas and Bruno Sansó (2016) "Bayesian Nonparametric Mode ling For Integro-Difference Equations". To appear in Statistics and Computing.