Are the Oceans Warming?
Statistical Models for Ocean Temperatures

Bruno Sansó

Department of Applied Mathematics and Statistics
University of California Santa Cruz
http://www.ams.ucsc.edu/∼bruno

In collaboration with Ricardo Lemos, Lisboa, Portugal.

Applied Mathematics and Statistics

- I am the Chair of the Department of Applied Mathematics and Statistics at UCSC. The Department is about eight years old.

- I am the Chair of the Department of Applied Mathematics and Statistics at UCSC. The Department is about eight years old.
- AMS is **NOT** the same as the Department of Mathematics at UCSC. We are part of the Jack Baskin School of Engineering.

Applied Mathematics and Statistics

- I am the Chair of the Department of Applied Mathematics and Statistics at UCSC. The Department is about eight years old.
- AMS is **NOT** the same as the Department of Mathematics at UCSC. We are part of the Jack Baskin School of Engineering.
- We offer MS and PhD in Statistics and Applied Mathematics. The application period is in the Fall of each year with and early January deadline.

Applied Mathematics and Statistics

- I am the Chair of the Department of Applied Mathematics and Statistics at UCSC. The Department is about eight years old.
- AMS is **NOT** the same as the Department of Mathematics at UCSC. We are part of the Jack Baskin School of Engineering.
- We offer MS and PhD in Statistics and Applied Mathematics. The application period is in the Fall of each year with and early January deadline.
- There are 12 faculty in AMS, six in Applied Mathematics and six in Statistics.

- Mathematical Biology: Marc Mangel and Hongyun Wang.

Applied Mathematics

- Mathematical Biology: Marc Mangel and Hongyun Wang.
- Astrophysics: Nic Brummell and Pascale Garaud.

- Mathematical Biology: Marc Mangel and Hongyun Wang.
- Astrophysics: Nic Brummell and Pascale Garaud.
- Control: Qi Gong and Dejan Milutinovic.

- Bayesian methods: all of us.

- Bayesian methods: all of us.
- Bayes Non-parametric: David draper, Athanasios Kottas and Abel Rodríguez.

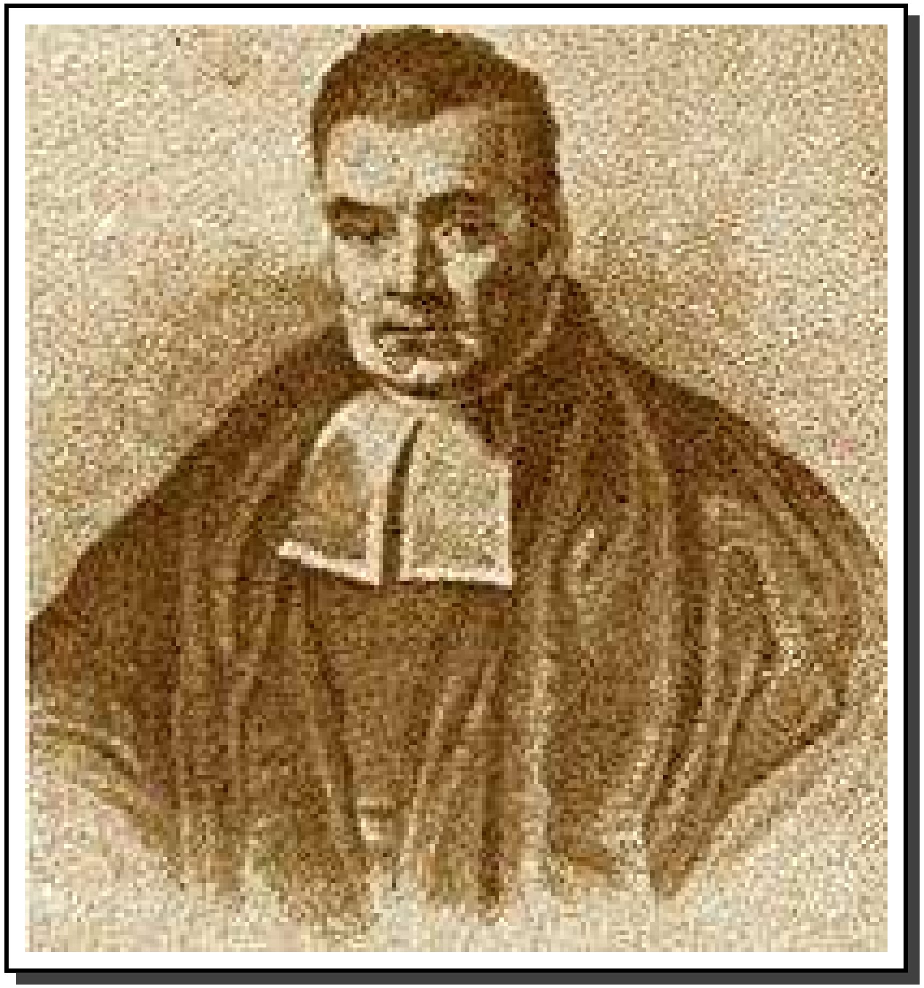
- Bayesian methods: all of us.
- Bayes Non-parametric: David draper, Athanasios Kottas and Abel Rodríguez.
- Computer models: Herbie Lee, Athanasios Kottas and Bruno Sansó.

- Bayesian methods: all of us.
- Bayes Non-parametric: David draper, Athanasios Kottas and Abel Rodríguez.
- Computer models: Herbie Lee, Athanasios Kottas and Bruno Sansó.
- Environmental applications: Athanasios Kottas and Bruno Sansó.

- Bayesian methods: all of us.
- Bayes Non-parametric: David draper, Athanasios Kottas and Abel Rodríguez.
- Computer models: Herbie Lee, Athanasios Kottas and Bruno Sansó.
- Environmental applications: Athanasios Kottas and Bruno Sansó.
- Time series: Abel Rodríguez, Raquel Prado.

- Bayesian methods: all of us.
- Bayes Non-parametric: David draper, Athanasios Kottas and Abel Rodríguez.
- Computer models: Herbie Lee, Athanasios Kottas and Bruno Sansó.
- Environmental applications: Athanasios Kottas and Bruno Sansó.
- Time series: Abel Rodríguez, Raquel Prado.
- Medical/biological applications: David Draper, Raquel Prado, Abel Rodríguez.

Thomas Bayes: 1702 – 1761



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Hierarchical Models

A useful framework for statistical models applied to problems in the physical sciences is given by hierarchical models with the levels:

Observational Model:

$$p(Y|\theta)$$

measurement error

Process Model:

$$p(\theta|\lambda)$$

process variability

Parameters Model:

$$p(\lambda)$$

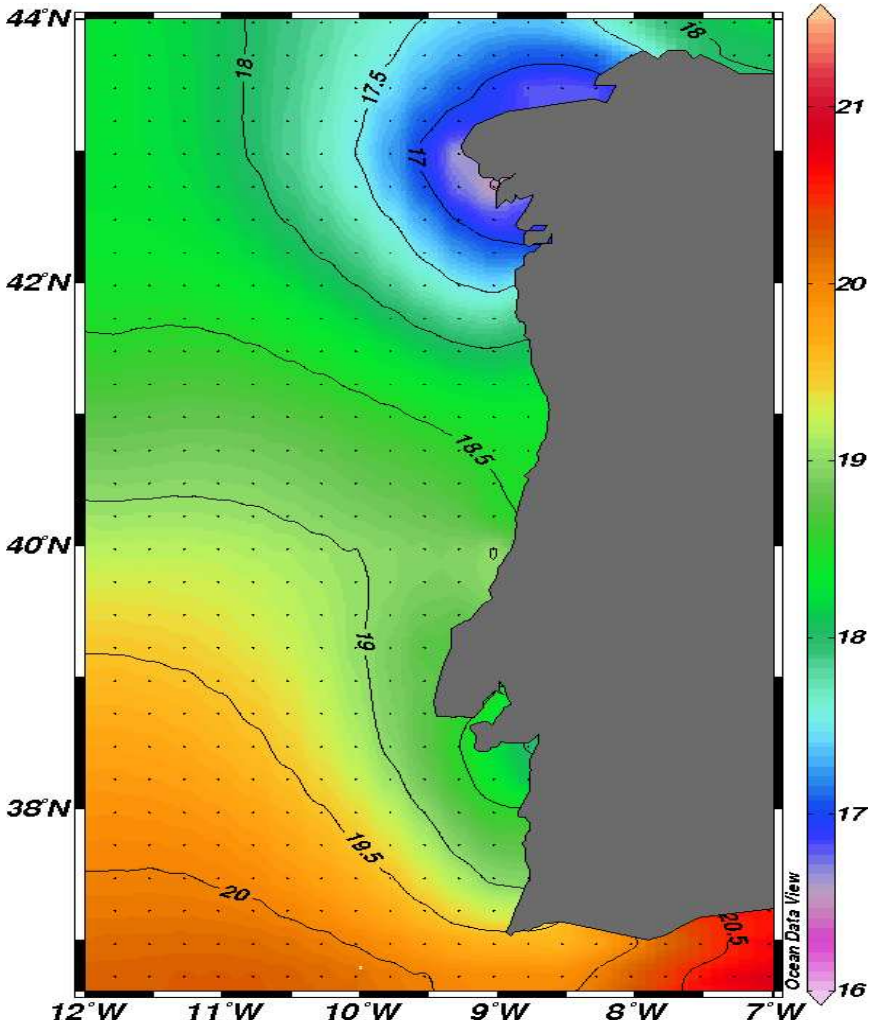
prior information

To learn about θ we obtain the posterior distribution $p(\theta|Y)$ using Bayes theorem.

Unfortunately, oftentimes the posterior distribution is intractable. A popular approach is to obtain samples from $p(\theta|Y)$ and use them for inference on θ.

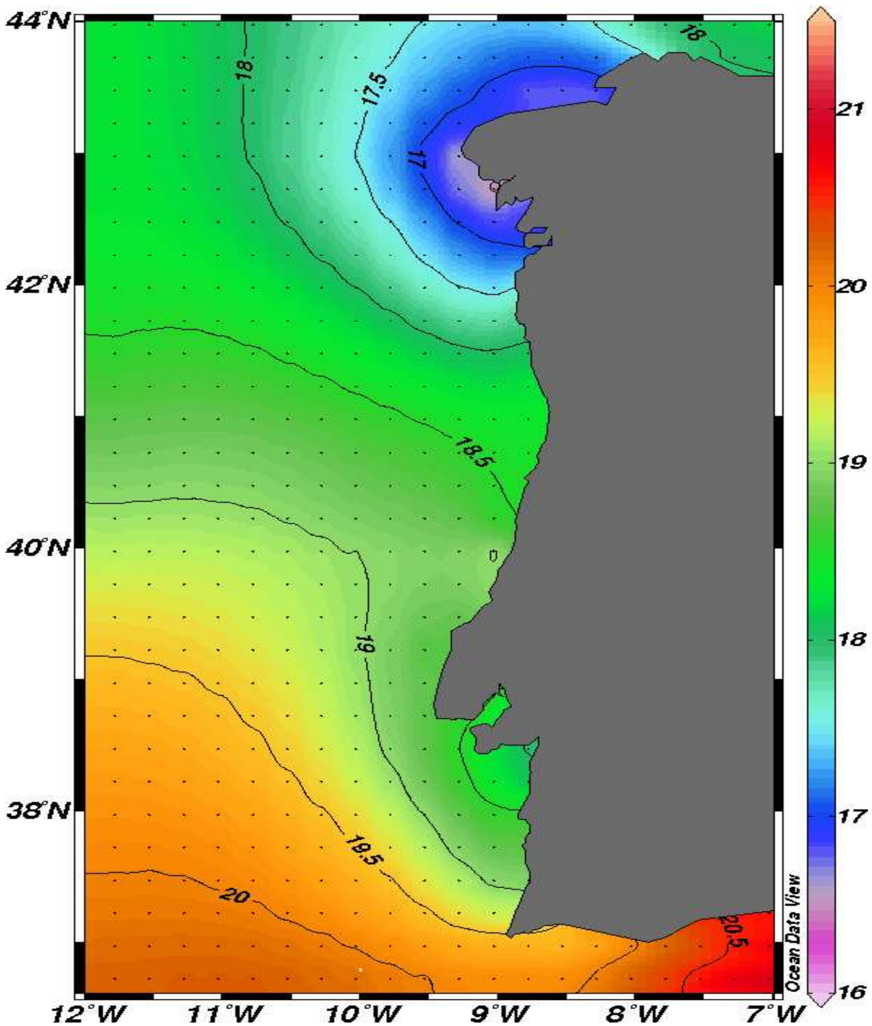
A very common method for sample-based inference is to simulate a Markov chain whose equilibrium distribution is $p(\theta|Y)$. After running the chain for a "burn in" period necessary to reach steady state the samples are collected and used for inference.

July average SST from WOA.



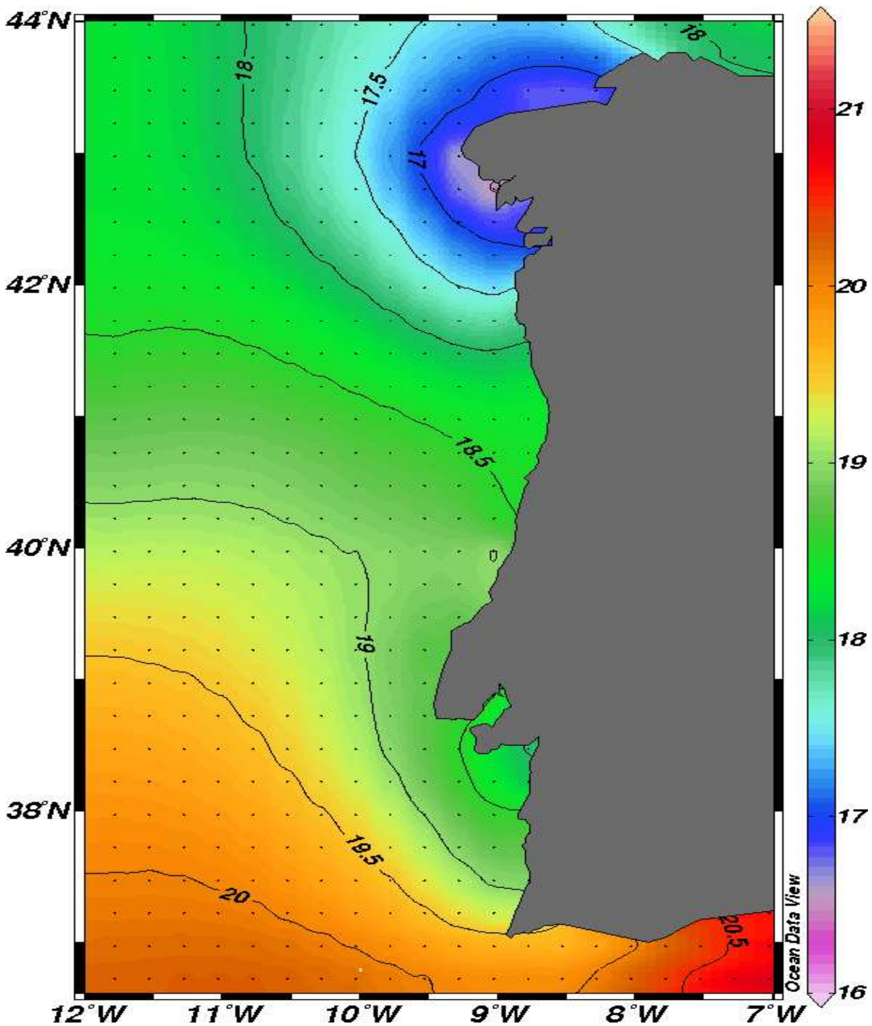
- Climatologies of ocean temperatures describe the mean state of the ocean using historical records.

July average SST from WOA.



- Climatologies of ocean temperatures describe the mean state of the ocean using historical records.
- They are used to understand the properties, distribution and circulation of water masses.

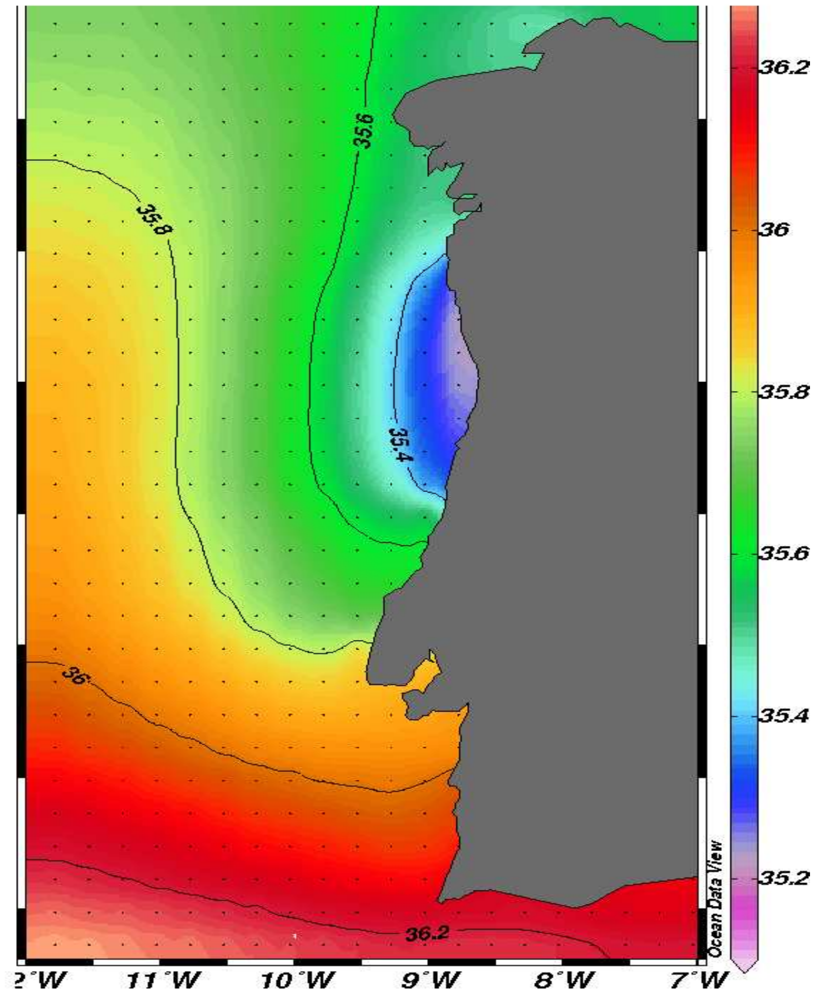
JULY AVERAGE SST FROM WOA.



- Climatologies of ocean temperatures describe the mean state of the ocean using historical records.
- They are used to understand the properties, distribution and circulation of water masses.
- They are also used to calibrate remote sensors and for the spin up, forcing, relaxation and validation of numerical models.

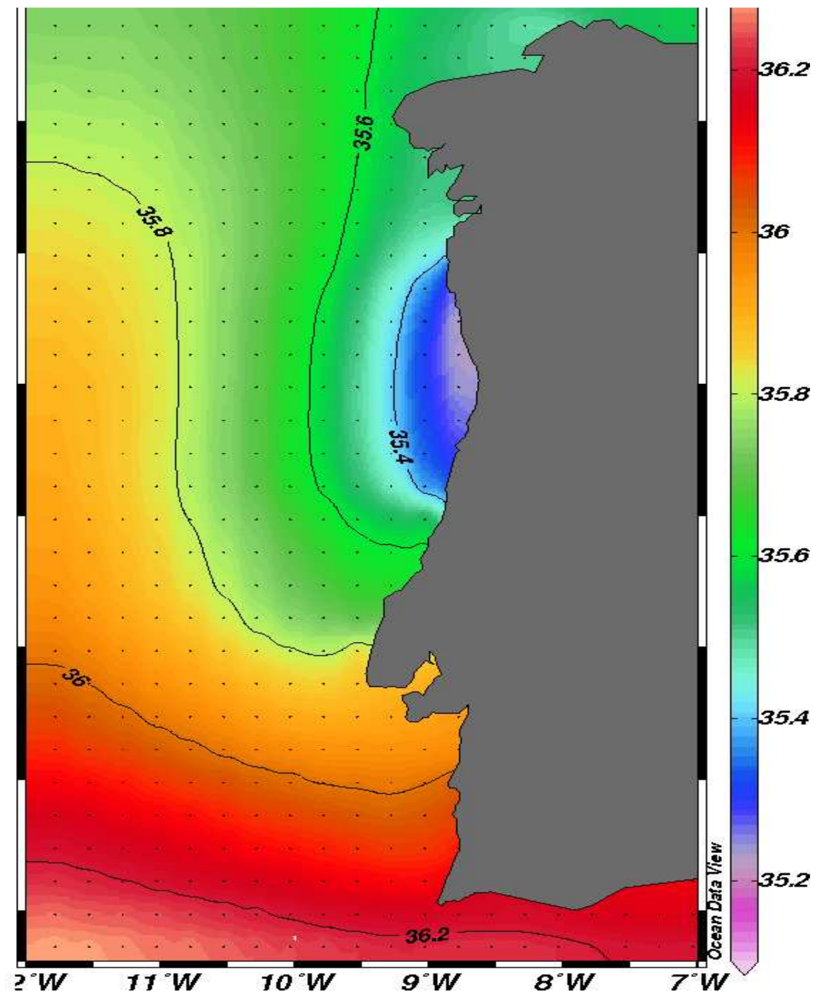
JULY AVERAGE SSS FROM WOA.

- In addition to the climatic mean, scientists are interested in the variability around the mean, or anomalies.



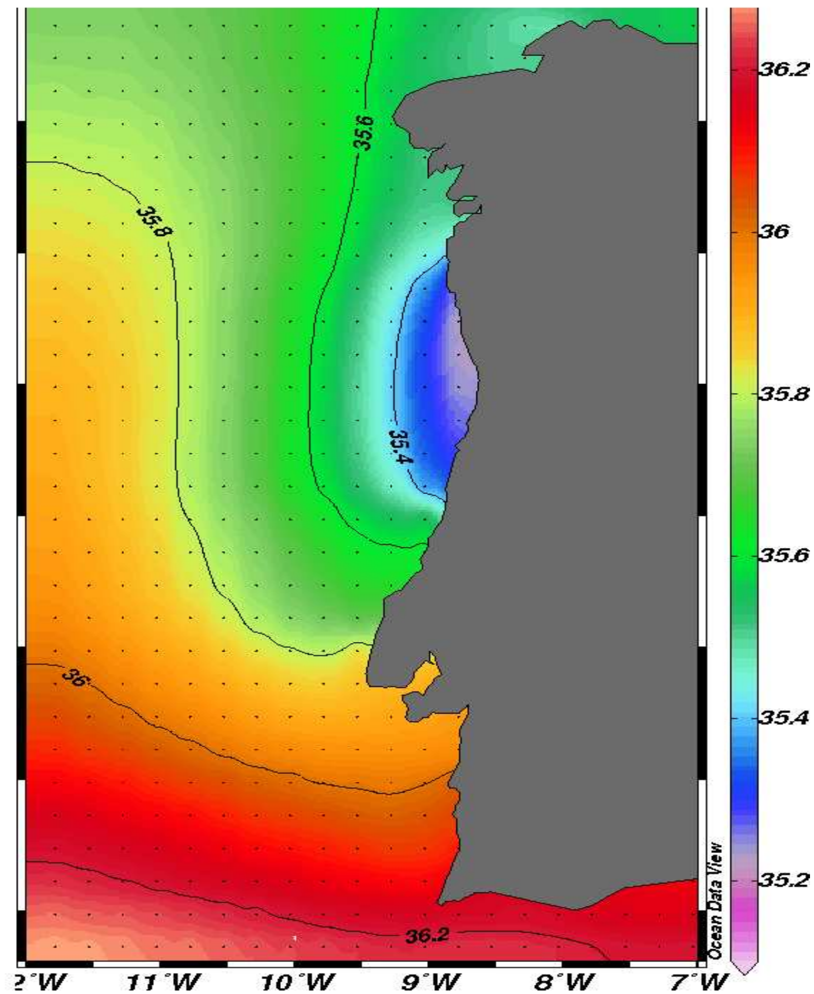
July average SSS from WOA.

- In addition to the climatic mean, scientists are interested in the variability around the mean, or anomalies.
- Finally, there is keen interest in the detection of long-term changes in ocean properties.



July average SSS from WOA.

- In addition to the climatic mean, scientists are interested in the variability around the mean, or anomalies.
- Finally, there is keen interest in the detection of long-term changes in ocean properties.
- At present, one of the standard climatological products is provided by the National Oceanographic Data Center (NODC) and is the World Ocean Atlas 2001, version 2.



The goals of our applications are the following:

The goals of our applications are the following:

• The climatology must be smooth and realistic, when compared to other existing products like the World Ocean Atlas (WOA).

Goals

The goals of our applications are the following:

• The climatology must be smooth and realistic, when compared to other existing products like the World Ocean Atlas (WOA).

• Anomaly fields must capture medium- to large-scale features and average zero everywhere

Goals

The goals of our applications are the following:

• The climatology must be smooth and realistic, when compared to other existing products like the World Ocean Atlas (WOA).

• Anomaly fields must capture medium- to large-scale features and average zero everywhere

• The trend fields must be smooth.

Goals

The goals of our applications are the following:

- The climatology must be smooth and realistic, when compared to other existing products like the World Ocean Atlas (WOA).
- Anomaly fields must capture medium- to large-scale features and average zero everywhere
- The trend fields must be smooth.
- All fields must be accompanied by measures of uncertainty.

The goals of our applications are the following:

• The climatology must be smooth and realistic, when compared to other existing products like the World Ocean Atlas (WOA).

• Anomaly fields must capture medium- to large-scale features and average zero everywhere

• The trend fields must be smooth.

• All fields must be accompanied by measures of uncertainty.

• The method must be useful in a large geographical domains and long time frames.

The goals of our applications are the following:

- The climatology must be smooth and realistic, when compared to other existing products like the World Ocean Atlas (WOA).
- Anomaly fields must capture medium- to large-scale features and average zero everywhere
- The trend fields must be smooth.
- All fields must be accompanied by measures of uncertainty.
- The method must be useful in a large geographical domains and long time frames.
- Observational errors should be accounted for.

Why a Space-Time Model?

Why use a sophisticated spatio-temporal model when an average is all that is needed?

Why a Space-Time Model?

Why use a sophisticated spatio-temporal model when an average is all that is needed?

- Data distribution is uneven in space and time. A model can borrow strength from data-dense periods and locations.

Why a Space-Time Model?

Why use a sophisticated spatio-temporal model when an average is all that is needed?

• Data distribution is uneven in space and time. A model can borrow strength from data-dense periods and locations.

• To account for location- and time-varying seasonal cycles, long term trends and high frequency variability.

Why a Space-Time Model?

Why use a sophisticated spatio-temporal model when an average is all that is needed?

• Data distribution is uneven in space and time. A model can borrow strength from data-dense periods and locations.

• To account for location- and time-varying seasonal cycles, long term trends and high frequency variability.

• To account for observational error.

Why a Space-Time Model?

Why use a sophisticated spatio-temporal model when an average is all that is needed?

• Data distribution is uneven in space and time. A model can borrow strength from data-dense periods and locations.

• To account for location- and time-varying seasonal cycles, long term trends and high frequency variability.

• To account for observational error.

• To incorporate different sources information, including structural knowledge.

Why a Space-Time Model?

Why use a sophisticated spatio-temporal model when an average is all that is needed?

• Data distribution is uneven in space and time. A model can borrow strength from data-dense periods and locations.

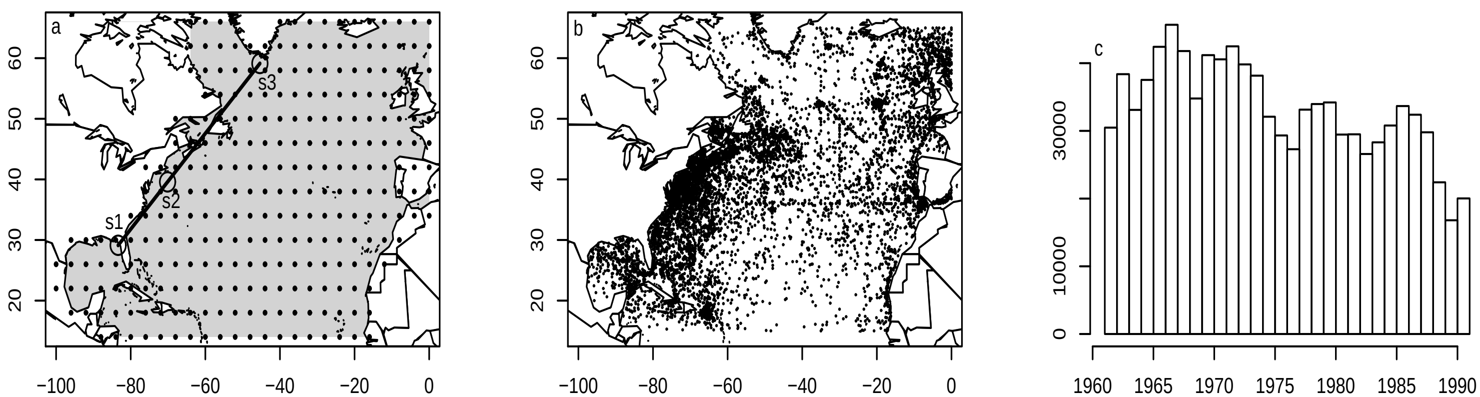
• To account for location- and time-varying seasonal cycles, long term trends and high frequency variability.

• To account for observational error.

• To incorporate different sources information, including structural knowledge.

• To produce probabilistic measures of uncertainty.

The domain S (gray area) and the grid J (bullets; $r_J = 4°$) used for the convolving process. A transect with three "case study" points. A random sample of 1% of the data. Temporal distribution of the data.



We use data from the NODC World Ocean Database 2005, collected with four types of instruments between 1961 and 1990.

We use data from the NODC World Ocean Database 2005, collected with four types of instruments between 1961 and 1990.

Data set name	no. obs.	variance
OSD	261,172	6.25×10^{-2}
CTD	29,879	6.25×10^{-6}
XBT	419,263	2.5×10^{-3}
MBT	439,783	2.025×10^{-1}

Discrete Process Convolutions

To reach the above goals, we build flexible spatio-temporal processes based on representing a Gaussian process as the convolution of simple processes over a grid.

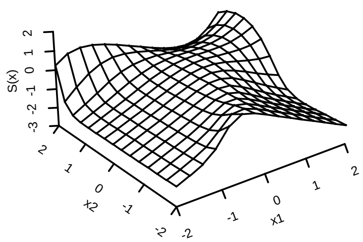
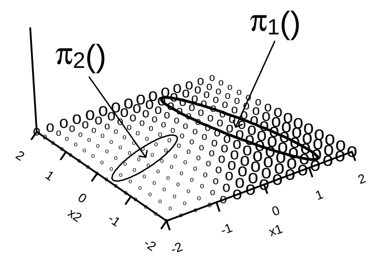
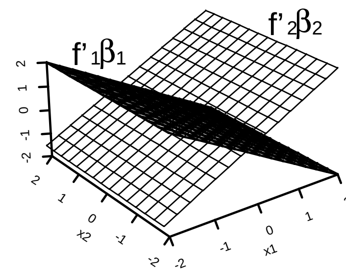
Discrete Process Convolutions

To reach the above goals, we build flexible spatio-temporal processes based on representing a Gaussian process as the convolution of simple processes over a grid.

For any point $\boldsymbol{s}$ in space S, say $\theta(\boldsymbol{s})$, we write

$$\theta(\boldsymbol{s}) = \sum_{\boldsymbol{j} \in J} K[\boldsymbol{s} - \boldsymbol{j}, \boldsymbol{\omega}(\boldsymbol{s})]\psi(\boldsymbol{j})$$

J is a grid in S. $K[\cdot, \boldsymbol{\omega}]$ is a kernel depending on a parameters $\boldsymbol{\omega}$, and $\psi(\boldsymbol{j})$ is a random field with a simple correlation structure.We term this a Discrete Process Convolution.

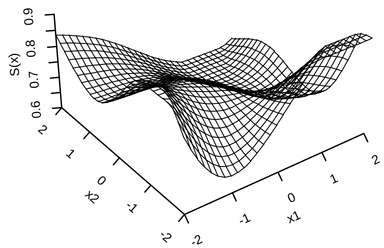
Spatial Process	Weighting Kernels	Component Surfaces
		

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

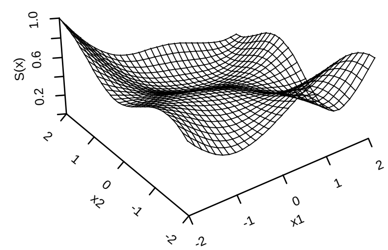
Spatial process generated from a mixture with $J = 2$. $\pi_j(\cdot)$ are weighting kernels. $\boldsymbol{f}_j'(s)\boldsymbol{\beta}_j$ are linear regression surfaces.

Dynamic Conditionally Linear Models

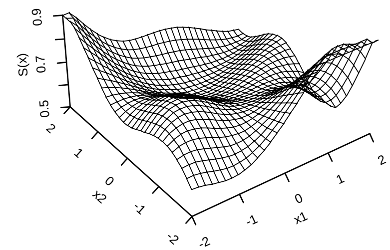
Time 1



Time 2



Time 3



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

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.

Model for North Atlantic SSTs

The SST observation $x_{i,m,y}(\boldsymbol{s})$ collected with instrument $i = 1, \ldots, 4$, in month m, year y and location $\boldsymbol{s}$ follows

$$x_{i,m,y}(\boldsymbol{s}) \sim N\left(\theta_{m,y}(\boldsymbol{s}), \tau_i^2\right).$$

Model for North Atlantic SSTs

The SST observation $x_{i,m,y}(\boldsymbol{s})$ collected with instrument $i = 1, \ldots, 4$, in month m, year y and location $\boldsymbol{s}$ follows

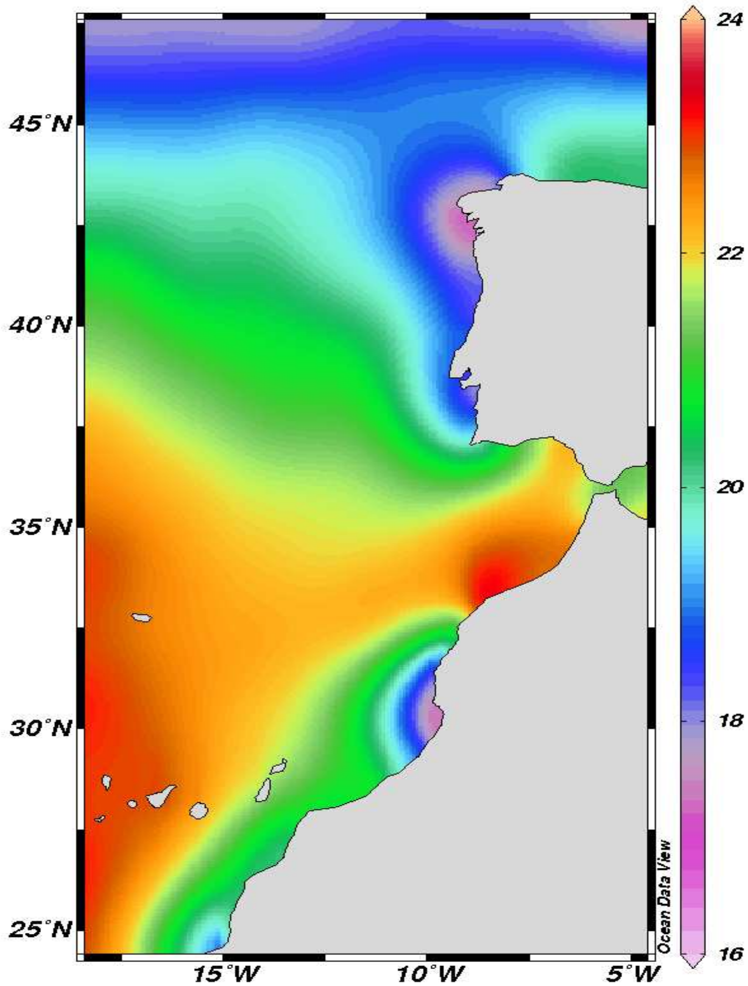
$$x_{i,m,y}(\boldsymbol{s}) \sim N\left(\theta_{m,y}(\boldsymbol{s}), \tau_i^2\right).$$

$$\theta_{m,y}(\boldsymbol{s}) \sim N\Big(\sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{\Lambda}(\boldsymbol{s})]\left(\alpha(\boldsymbol{j}) + \boldsymbol{\beta}_t(\boldsymbol{j})\boldsymbol{w}_t^T + \eta(\boldsymbol{j})(t-180)\right), \Phi(\boldsymbol{s})^2\Big).$$

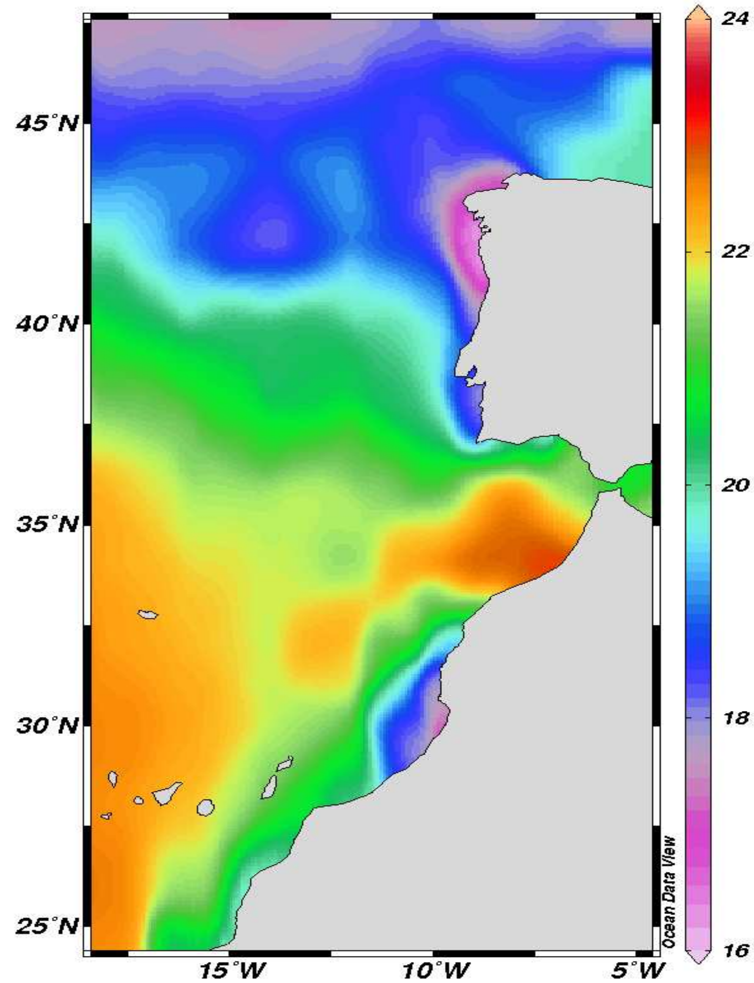
$$t = m + 12(y-1961) \quad \text{and} \quad \Phi(\boldsymbol{s})^2 = \sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{\Omega}(\boldsymbol{s})] \exp\left(\sigma(\boldsymbol{j})\right).$$

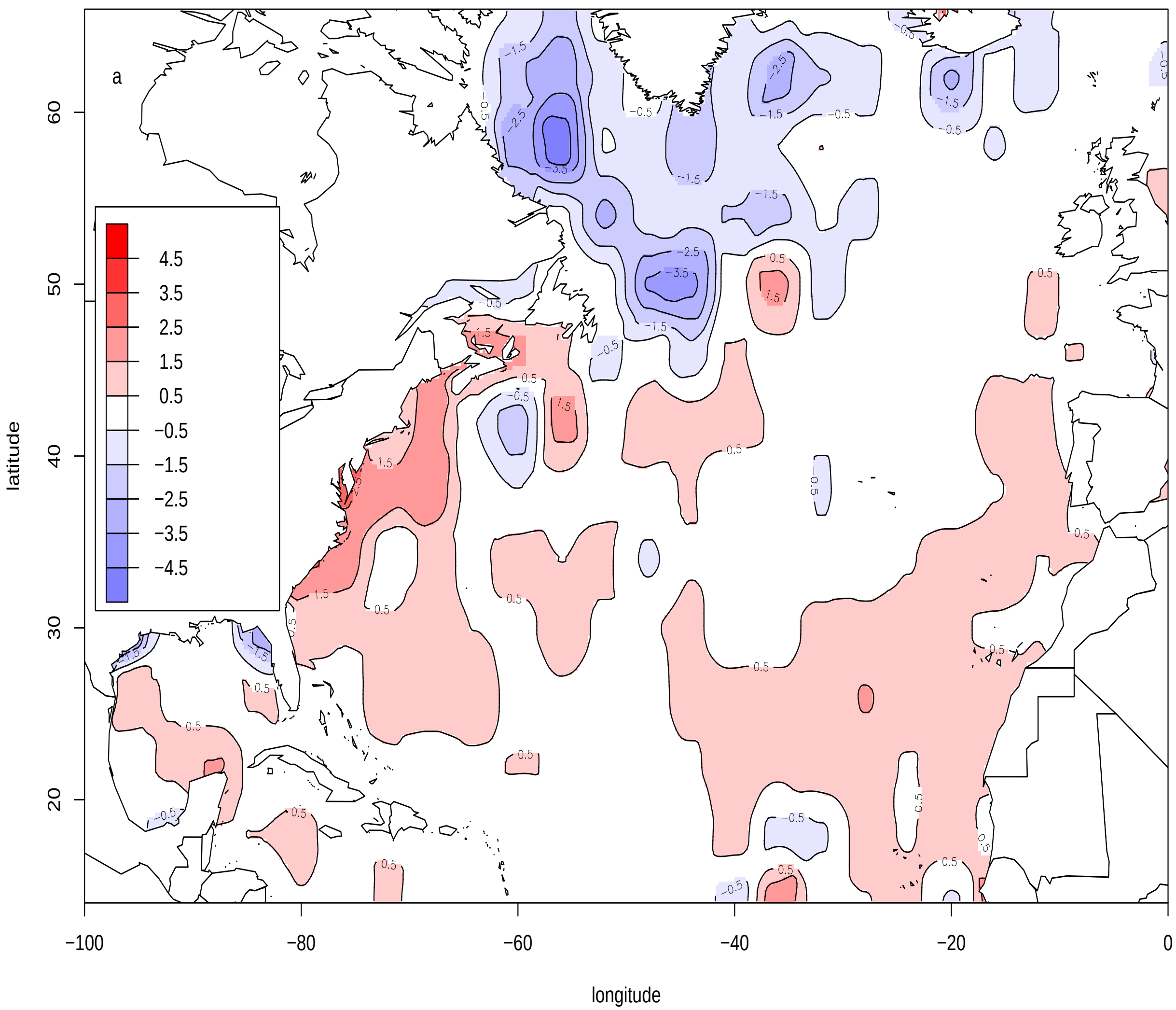
Results: August Climatology

August SST from WOA01



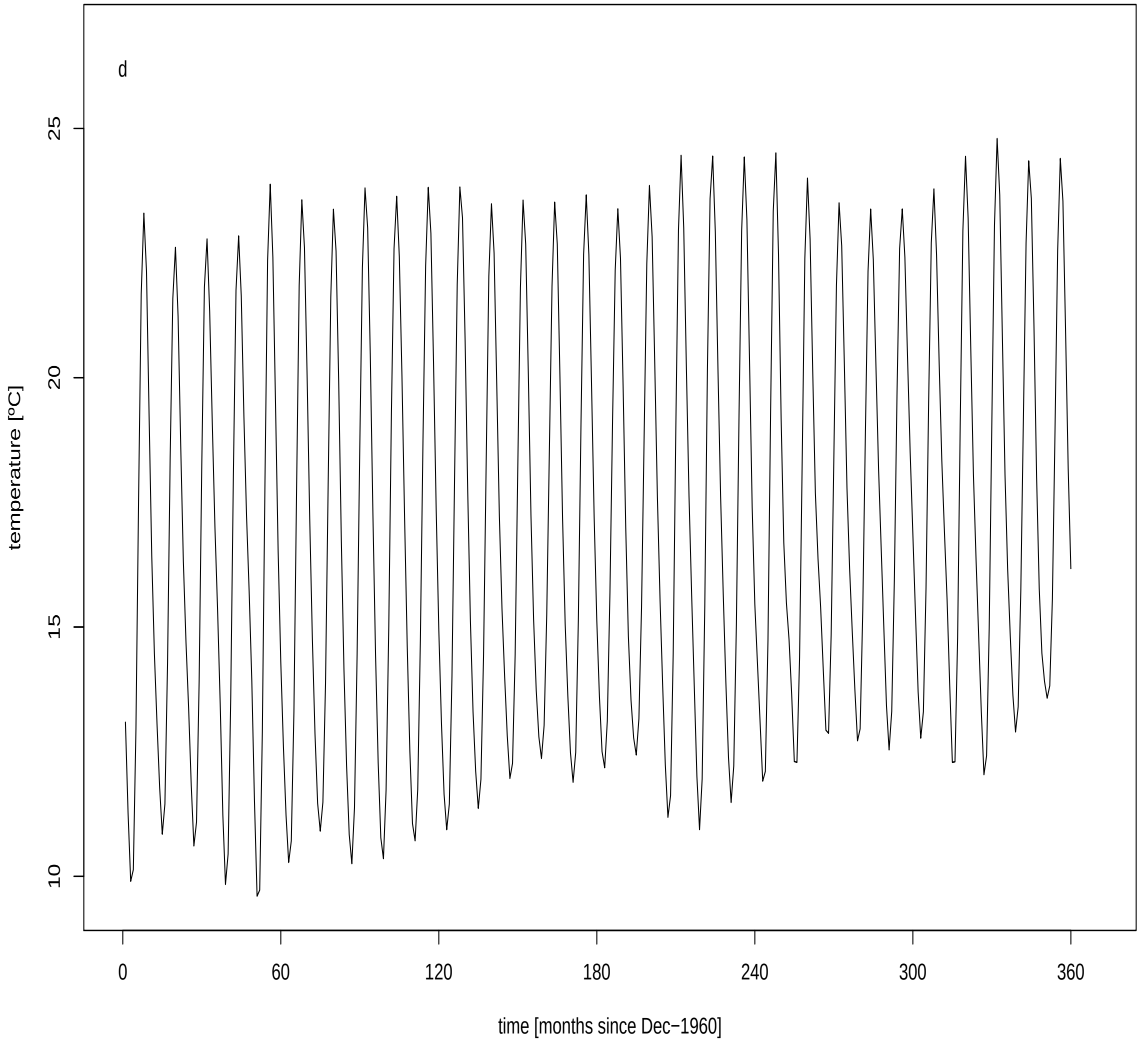
August SST from LS09





Trends (°C/30 years)

Posterior mean for monthly SST at s_2 (°C).



Model for Salinity and Temperature

The development of a model that considers temperature and salinity jointly over a 3D domain needs to account the following issues:

Model for Salinity and Temperature

The development of a model that considers temperature and salinity jointly over a 3D domain needs to account the following issues:

• Data sparsity. There are substantially more observations for the surface of the ocean than for the deep water. Furthermore, salinity is sampled about 10 times less than temperature.

Model for Salinity and Temperature

The development of a model that considers temperature and salinity jointly over a 3D domain needs to account the following issues:

• Data sparsity. There are substantially more observations for the surface of the ocean than for the deep water. Furthermore, salinity is sampled about 10 times less than temperature.

• The column of water needs to satisfy a density stability constraint. That is, density must increase with depth. Density is related to temperature and salinity via the equation of state.

Model for Salinity and Temperature

The development of a model that considers temperature and salinity jointly over a 3D domain needs to account the following issues:

• Data sparsity. There are substantially more observations for the surface of the ocean than for the deep water. Furthermore, salinity is sampled about 10 times less than temperature.

• The column of water needs to satisfy a density stability constraint. That is, density must increase with depth. Density is related to temperature and salinity via the equation of state.

• The previous two points imply that we need to consider a hierarchical structure for our statistical model that is inspired by physical equations.

We consider observations averaged over grid cells indexed by $\boldsymbol{i} = (i_1, i_2, i_3) \in \mathbb{I}$. The observation equation is given by

$$\begin{pmatrix} T_t(\boldsymbol{i}) \\ S_t(\boldsymbol{i}) \end{pmatrix} \sim N_2 \left(\boldsymbol{\theta}_t(\boldsymbol{i}), \begin{pmatrix} \frac{\tau_T^2}{n_{T,t}(\boldsymbol{i})} & 0 \\ 0 & \frac{\tau_S^2}{n_{S,t}(\boldsymbol{i})} \end{pmatrix} \right)$$

We consider observations averaged over grid cells indexed by $\boldsymbol{i} = (i_1, i_2, i_3) \in \mathbb{I}$. The observation equation is given by

$$\begin{pmatrix} T_t(\boldsymbol{i}) \\ S_t(\boldsymbol{i}) \end{pmatrix} \sim N_2\left(\boldsymbol{\theta}_t(\boldsymbol{i}), \begin{pmatrix} \frac{\tau_T^2}{n_{T,t}(\boldsymbol{i})} & 0 \\ 0 & \frac{\tau_S^2}{n_{S,t}(\boldsymbol{i})} \end{pmatrix}\right)$$

The process equation is

$$\boldsymbol{\theta}_t(\boldsymbol{i}) \sim N_2\left(\sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{i}-\boldsymbol{j}, \boldsymbol{w}(\boldsymbol{i})]\boldsymbol{\lambda}_t(\boldsymbol{j}), \boldsymbol{C}(\boldsymbol{i})\right)$$

$$\boldsymbol{C}(\boldsymbol{i}) = \tau_{\theta_1}\tau_{\theta_2}\begin{pmatrix} \frac{\tau_{\theta_1}}{\tau_{\theta_2}} & \Phi\left(\sum_{\boldsymbol{j}} K[\boldsymbol{i}-\boldsymbol{j}]\varpi(\boldsymbol{j})\right) \\ \Phi\left(\sum_{\boldsymbol{j}} K[\boldsymbol{i}-\boldsymbol{j}]\varpi(\boldsymbol{j})\right) & \frac{\tau_{\theta_2}}{\tau_{\theta_1}} \end{pmatrix}$$

with $\boldsymbol{j}$ in a grid $\mathbb{J}$, $\boldsymbol{w}$ a DPC and an MRF for ϖ.

The Model for the Latent Process

$$\boldsymbol{\lambda}_t(\boldsymbol{j}) = \boldsymbol{\lambda}_{t-1}(\boldsymbol{j}) + \boldsymbol{\alpha}_{t-1}(\boldsymbol{j})s_t + \boldsymbol{\beta}_{t-1}(\boldsymbol{j})c_t + \boldsymbol{\gamma}_{t-1}(\boldsymbol{j}) + \boldsymbol{\epsilon}_t^{\lambda}(\boldsymbol{j})$$

$$\boldsymbol{\alpha}_t(\boldsymbol{j}) = \boldsymbol{\alpha}_{t-1}(\boldsymbol{j}) + \boldsymbol{\epsilon}_t^{\alpha}(\boldsymbol{j})$$

$$\boldsymbol{\beta}_t(\boldsymbol{j}) = \boldsymbol{\beta}_{t-1}(\boldsymbol{j}) + \boldsymbol{\epsilon}_t^{\beta}(\boldsymbol{j})$$

$$\boldsymbol{\gamma}_t(\boldsymbol{j}) = \boldsymbol{\gamma}_{t-1}(\boldsymbol{j}) + \boldsymbol{\epsilon}_t^{\gamma}(\boldsymbol{j})$$

with $(\boldsymbol{\epsilon}_t^{\lambda}(\boldsymbol{j}), \boldsymbol{\epsilon}_t^{\alpha}(\boldsymbol{j}), \boldsymbol{\epsilon}_t^{\beta}(\boldsymbol{j}), \boldsymbol{\epsilon}_t^{\gamma}(\boldsymbol{j}))' \sim N_8\left(\mathbf{0}, \boldsymbol{W}_t(\boldsymbol{j})\right)$, $\boldsymbol{W}_t(\boldsymbol{j})$ obtained using discount factors, $s_t = \sin\left(2\pi t/52\right)$ and $c_t = \cos\left(2\pi t/52\right)$.

The Model for Velocities

The physical constraints are imposed by assuming dependence on the zonal and meridional velocities, $\upsilon_t(\boldsymbol{i})$ respectively, and $\nu_t(\boldsymbol{i})$ and the density $\rho_t(\boldsymbol{i})$.

$$\upsilon_t(\boldsymbol{i}) \sim N^{(\upsilon_{\min}(\boldsymbol{i}),\upsilon_{max}(\boldsymbol{i}))}\left(m_\upsilon, \tau_\upsilon^2\right)$$

$$\nu_t(\boldsymbol{i}) \sim N^{(\nu_{\min}(\boldsymbol{i}),\nu_{max}(\boldsymbol{i}))}\left(m_\nu, \tau_\nu^2\right)$$

where, assuming geostrophic conditions,

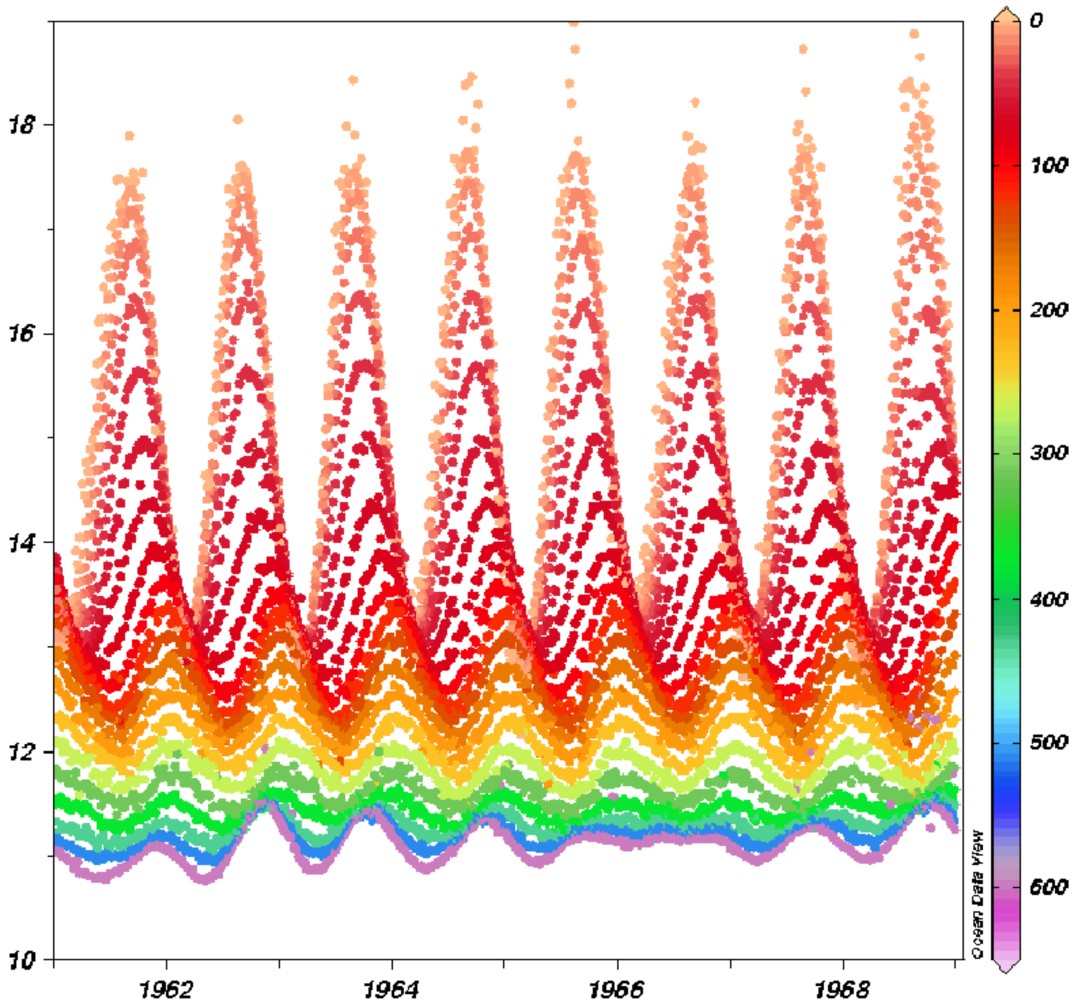
$$m_\upsilon = \frac{z(\boldsymbol{i})g(\boldsymbol{i})}{2\rho_t(\boldsymbol{i})f_c(\boldsymbol{i})}\frac{\partial\rho_t(\boldsymbol{i})}{\partial y} \quad \text{and} \quad m_\nu = -\frac{z(\boldsymbol{i})g(\boldsymbol{i})}{2\rho_t(\boldsymbol{i})f_c(\boldsymbol{i})}\frac{\partial\rho_t(\boldsymbol{i})}{\partial x}$$

Here $z(\boldsymbol{i})$ is thickness, $g(\boldsymbol{i})$ is the gravity acceleration and $f_c(\boldsymbol{i})$ is the Coriolis parameter.

Density is determined by the equation of state, $f_w(\cdot)$ that provides density as a function of temperature, salinity and pressure (depth). Thus

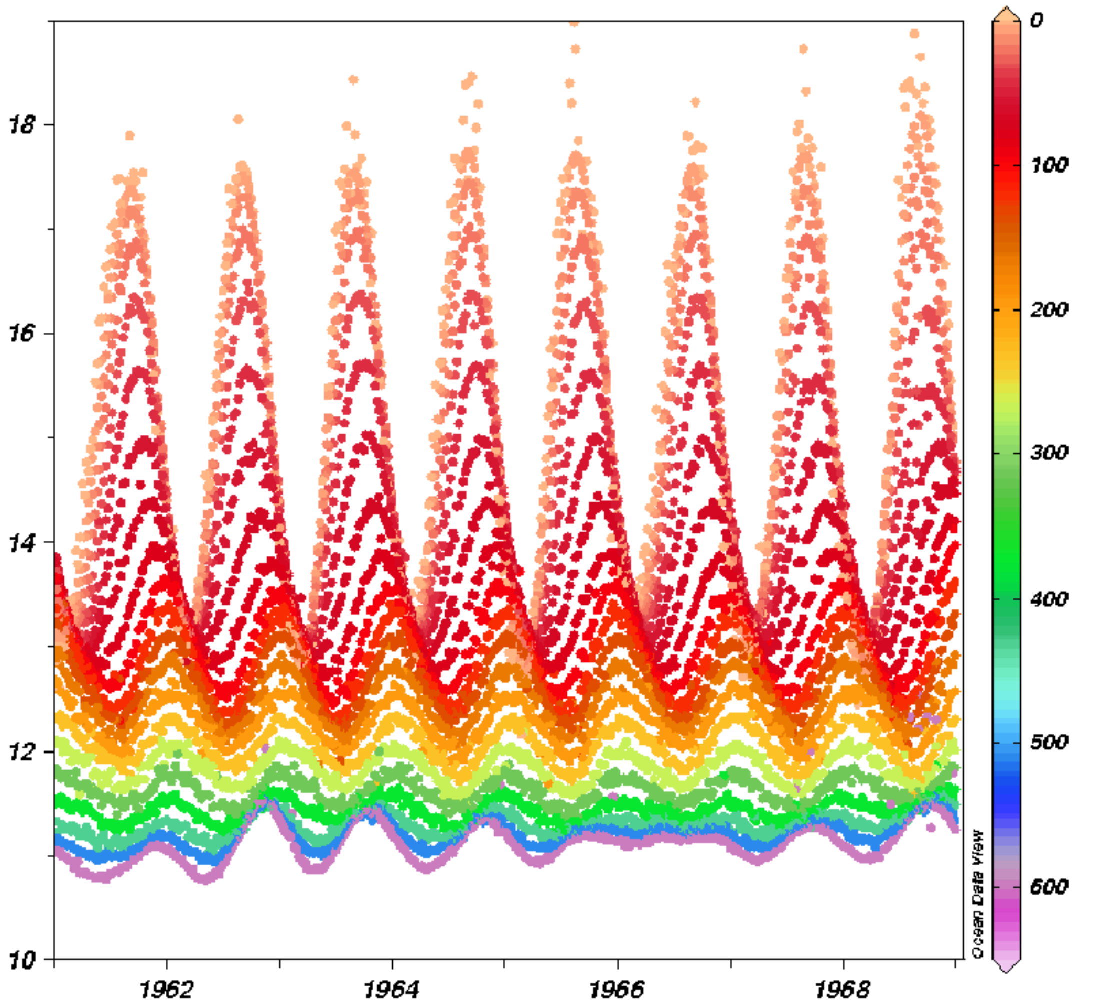
$$\rho_t(\boldsymbol{i}) \sim N^{(\rho_t(\mathcal{A}(\boldsymbol{i})),\rho_t(\mathcal{B}(\boldsymbol{i})))}\left(f_w\left(\boldsymbol{\theta}_t(\boldsymbol{i}), p(\boldsymbol{i})\right), \tau_\rho^2\right)$$

$\mathcal{A}(\boldsymbol{i})$ and $\mathcal{B}(\boldsymbol{i})$ denote respectively the grid points of $\mathbb{I}$ immediately above and below $\boldsymbol{i}$.

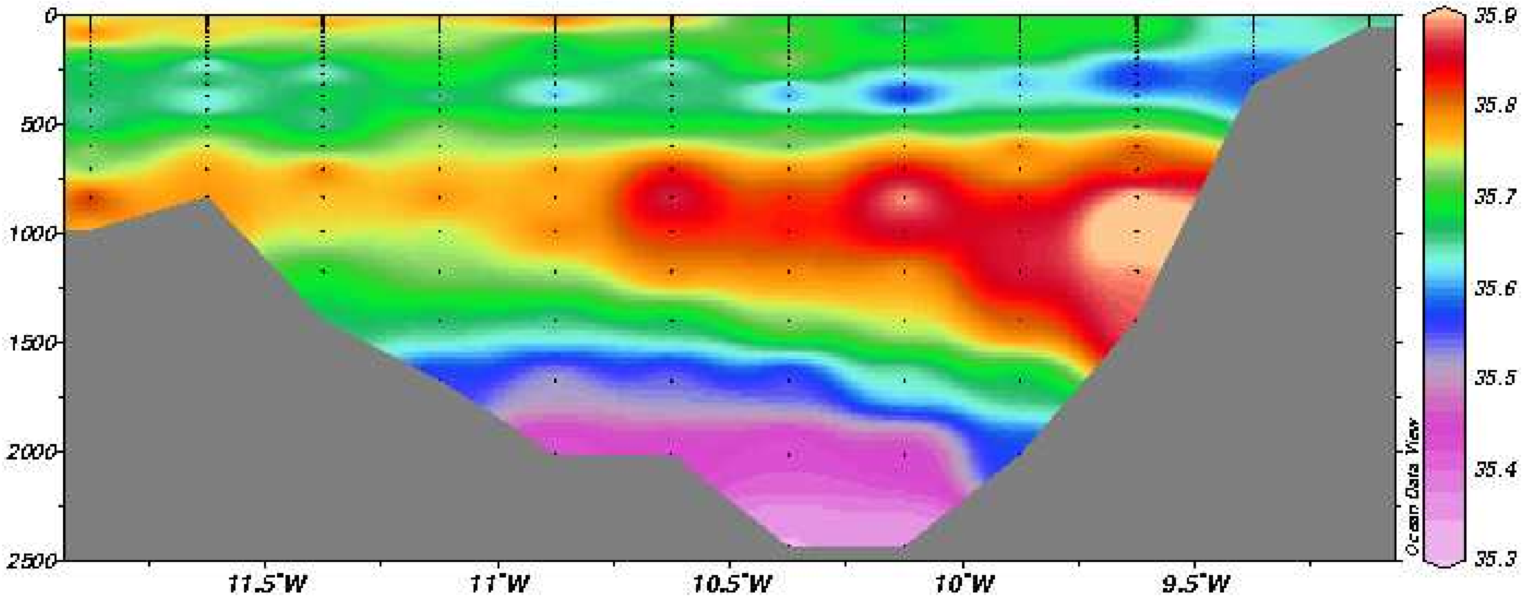


- We run the 3D joint model for weekly data corresponding to the Iberian Peninsula from 1961 to 1970.

Temperature Variability

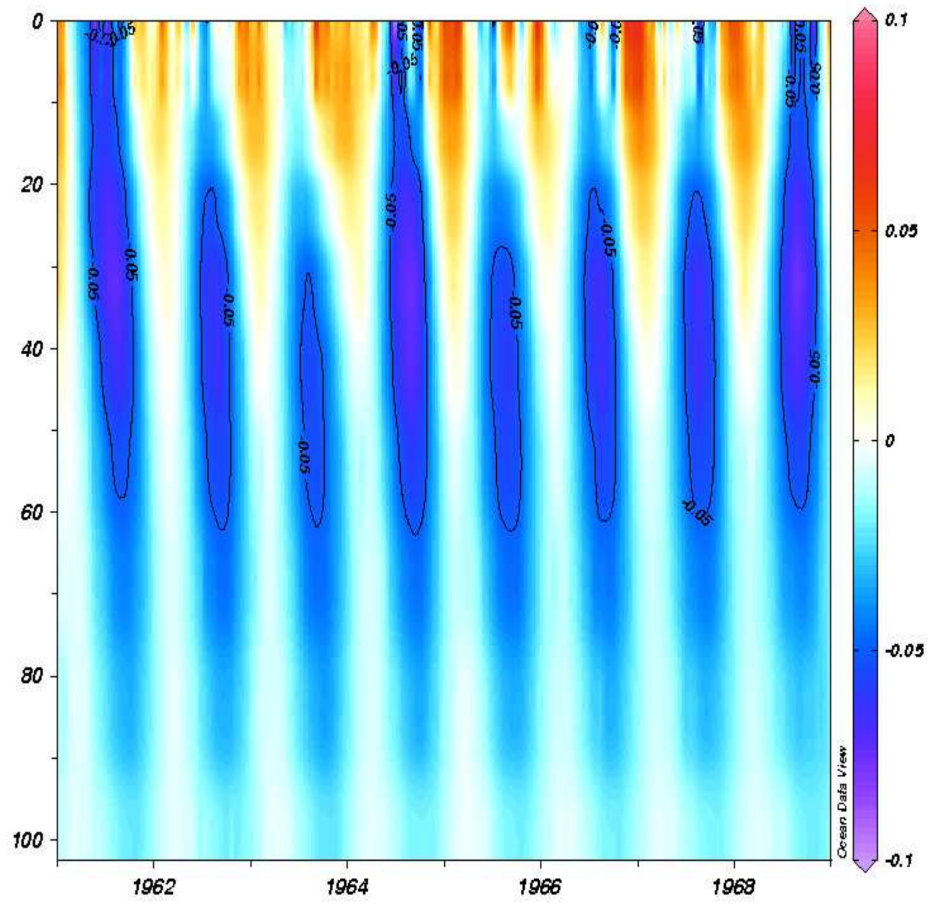


- We run the 3D joint model for weekly data corresponding to the Iberian Peninsula from 1961 to 1970.
- The figure shows the estimated temperatures for different depths in an area off the coast of Galicia.



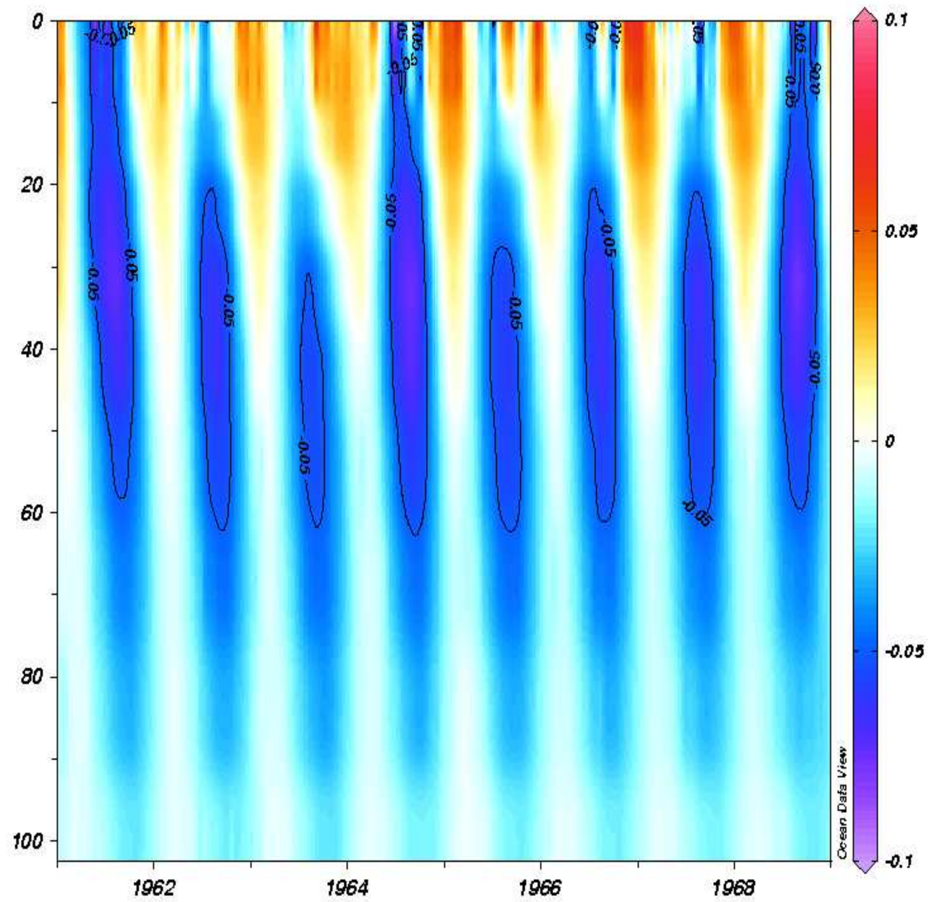
Estimated salinity field for an area off the coast of Galicia in July 1961. We observe a very clear plume of salty water close to the coast that is known to come from the Mediterranean Sea.

Time varying thermocline at Cape St. Vincent.



- The Thermocline is a thin ocean layer where the temperature gradient is high relative to the upper and lower layers.

Time varying thermocline at Cape St. Vincent.



- The Thermocline is a thin ocean layer where the temperature gradient is high relative to the upper and lower layers.
- It separates, in terms of temperature, the high varying surface waters to the smoothly varying deep ocean waters.

Discussion

- We have developed spatio-temporal models for ocean variables that are non-stationary in either space or time. We are able to capture important space- and time-varying features of the ocean.

- We have developed spatio-temporal models for ocean variables that are non-stationary in either space or time. We are able to capture important space- and time-varying features of the ocean.
- Our model includes observational errors and realistic descriptions of the latent processes governing the evolution of ocean variables.

- We have developed spatio-temporal models for ocean variables that are non-stationary in either space or time. We are able to capture important space- and time-varying features of the ocean.
- Our model includes observational errors and realistic descriptions of the latent processes governing the evolution of ocean variables.
- The model is able to provide probabilistic assessments of the variabilities included in the estimated quantities. All estimation variabilities are accounted for in the final product.

- Our model is able to handle large data sets. By using kernels with compact support and making use of the structure of the CDLM we are able to parallelize the estimation algorithms.

Discussion

- Our model is able to handle large data sets. By using kernels with compact support and making use of the structure of the CDLM we are able to parallelize the estimation algorithms.

- We are currently refining the evolution equation to include advection and conservation of mass for the latent processes.

- Our model is able to handle large data sets. By using kernels with compact support and making use of the structure of the CDLM we are able to parallelize the estimation algorithms.

- We are currently refining the evolution equation to include advection and conservation of mass for the latent processes.

- Reference: Ricardo T. Lemos, Bruno Sansó (2006) "Spatio-temporal Variability of Ocean Temperature in the Portugal Current System". *Journal of Geophysical Research Oceans*, 111, C04010, doi:10.1029/2005JC003051.

- Our model is able to handle large data sets. By using kernels with compact support and making use of the structure of the CDLM we are able to parallelize the estimation algorithms.

- We are currently refining the evolution equation to include advection and conservation of mass for the latent processes.

- Reference: Ricardo T. Lemos, Bruno Sansó (2006) "Spatio-temporal Variability of Ocean Temperature in the Portugal Current System". *Journal of Geophysical Research Oceans*, 111, C04010, doi:10.1029/2005JC003051.

- Reference: Ricardo T. Lemos, Bruno Sansó (2009) "A Spatio-Temporal Model for Mean, Anomaly and Trend Fields of North Atlantic Sea Surface Temperature". *Journal of the American Statistical Association*, 104, pp. 5–25.