SPATIO-TEMPORAL STATISTICAL MODELLING OF SIGNIFICANT WAVE HEIGHT

Anastassia Baxevani
Chalmers University of Technology

Joint work with I. Rychlik and S. Caires

Introductory remarks I

Two sources of random variation are present when studying the sea surface:

- short-term scale or sea-state, which corresponds to stationary conditions on the sea

Introductory remarks I

Two sources of random variation are present when studying the sea surface:

- short-term scale or sea-state, which corresponds to stationary conditions on the sea
- long-term scale, which corresponds to evolution of functionals (H_s - significant wave height) defined on the sea states over long periods of time and at large geographical regions

Introductory remarks I

Two sources of random variation are present when studying the sea surface:

- short-term scale or sea-state, which corresponds to stationary conditions on the sea
- long-term scale, which corresponds to evolution of functionals (H_s - significant wave height) defined on the sea states over long periods of time and at large geographical regions
- H_s equals four times the standard deviation of the vertical displacement of the ocean surface $\rightarrow$ H_s^2 is proportional to the wave energy.

Introductory remarks II

- It has been reported in Anderson et al. 2001, that at fixed location $\log(H_s)$ are well approximated by means of AR processes.

Introductory remarks II

- It has been reported in Anderson et al. 2001, that at fixed location $\log(H_s)$ are well approximated by means of AR processes.
- It has been shown in Baxevani et al. 2006, that for fixed time over squares with sides $\approx 4^\circ$, $\log(H_s)$ can be modelled by means of homogeneous Gaussian fields.

Introductory remarks II

- It has been reported in Anderson et al. 2001, that at fixed location $\log(H_s)$ are well approximated by means of AR processes.
- It has been shown in Baxevani et al. 2006, that for fixed time over squares with sides $\approx 4^{\circ}$, $\log(H_s)$ can be modelled by means of homogeneous Gaussian fields.
- Our goal is to propose a global spatio-temporal model for the variation of H_s.

Introductory remarks II

- It has been reported in Anderson et al. 2001, that at fixed location $\log(H_s)$ are well approximated by means of AR processes.
- It has been shown in Baxevani et al. 2006, that for fixed time over squares with sides $\approx 4^\circ$, $\log(H_s)$ can be modelled by means of homogeneous Gaussian fields.
- Our goal is to propose a global spatio-temporal model for the variation of H_s.
- Data used consist of records of H_s measurements estimated by:
  - data from the TOPEX-Poseidon satellite, from buoys, C-ERA 40 data.

Plan

To use the spatial stationary models that are valid locally (over small regions of stationarity) and the information on the temporal models

$\Updownarrow$

to construct a non-homogeneous spatio-temporal model

Homogeneous model, spatial field

For fixed $t_0 \in \mathbb{R}^+$ let $\forall\ \mathbf{p}_0 \in \mathbb{R}^2, \eta^{t_0}(\mathbf{p}_0) \equiv \eta(\mathbf{p}_0) \subset \mathbb{R}^2$ denote the region of stationarity and $X(\mathbf{p}),\ \mathbf{p} = (p_1, p_2) \in \eta(\mathbf{p}_0)$ be a zero-mean stationary Gaussian field

- with covariance function

$$r(\mathbf{p}) = \text{Cov}(X(\mathbf{0}), X(\mathbf{p})) = \sigma^2 e^{-\frac{|\mathbf{p}|^2}{2L^2}},$$

where σ^2 is the field variance, $|\cdot| : \mathbb{R}^2 \to \mathbb{R}$ denotes the euclidean distance and L is the so-called correlation length

Homogeneous model, spatial field

For fixed $t_0 \in \mathbb{R}^+$ let $\forall\ \mathbf{p}_0 \in \mathbb{R}^2, \eta^{t_0}(\mathbf{p}_0) \equiv \eta(\mathbf{p}_0) \subset \mathbb{R}^2$ denote the region of stationarity and $X(\mathbf{p}),\ \mathbf{p} = (p_1, p_2) \in \eta(\mathbf{p}_0)$ be a zero-mean stationary Gaussian field

- with covariance function

$$r(\mathbf{p}) = \text{Cov}(X(\mathbf{0}), X(\mathbf{p})) = \sigma^2 e^{-\frac{|\mathbf{p}|^2}{2L^2}},$$

  where σ^2 is the field variance, $|\cdot| : \mathbb{R}^2 \to \mathbb{R}$ denotes the euclidean distance and L is the so-called correlation length

- and corresponding spectral density

$$S(\boldsymbol{\omega}) = S(\boldsymbol{\omega}; \sigma^2, L) = \frac{\sigma^2 L^2}{2\pi} e^{-L^2 \frac{|\boldsymbol{\omega}|^2}{2}}, \quad \boldsymbol{\omega} = (\omega_1, \omega_2) \in \mathbb{R}^2.$$

Homogeneous model, temporal field

For fixed $\mathbf{p} \in \eta^{t_0}(\mathbf{p}_0)$, let $X(t)$ be temporal field with

$$r(t;T) = \mathrm{Cov}(X(0), X(t)) = \sigma^2 e^{-\frac{t^2}{2T^2} - \frac{|t|}{2C}},$$

where

- T, C depend on $t_0, \mathbf{p}_0$
- $|\mathbf{v}| = \frac{L}{T}$, $\mathbf{v}$ is the velocity $X(\mathbf{p})$ is drifting with
- $r(t;\infty) = \sigma^2 e^{-\frac{|t|}{2C}}$ – i.e. no dynamics the field is Markov
- $C = \infty \Rightarrow r(t;T) = r(-\mathbf{v}t) = \mathrm{Cov}(X(\mathbf{p}), X(\mathbf{p} - \mathbf{v}t))$ – time variability is a consequence of the field drifting with constant velocity $\mathbf{v}$.

Homogeneous model, spatio-temporal field

Let $X(\mathbf{p}, t)$ be the field for $t, \mathbf{p} \in \eta(\mathbf{p}_0)$. Construct covariance

$$r(\mathbf{p}, t) = \text{Cov}(X(\mathbf{0}, 0), X(\mathbf{p}, t)) = \sigma^2 e^{-\frac{|t|}{2C} - \frac{|\mathbf{p}-\mathbf{v}t|^2}{2L^2}} = \rho(t) r(\mathbf{p}-\mathbf{v}t).$$

For any $t = i\,dt,\ i > 0,\ dt > 0$ let

$$X(\mathbf{p}, t) = \rho^i X_0^*(\mathbf{p} - \mathbf{v}t) + \sum_{k=1}^{i} \rho^{i-k} \sqrt{1-\rho^2} X_k^*(\mathbf{p} - \mathbf{v}(i-k)\,dt)$$

with $X(\mathbf{p}, 0) = X_0^*(\mathbf{p})$ and $X_i^*(\mathbf{p})$ be homogeneous, i.i.d r.fs with spatial covariance $r(\mathbf{p})$ and $\rho = \rho(\,dt) = e^{-\frac{dt}{2C}}$.

Non homogeneous model, spatial field

Let $L(\mathbf{p})$ and $\sigma^2(\mathbf{p})$ be slowly varying, and $X(\mathbf{p})$ be a Gaussian field with spatial cov. for $\mathbf{p}, \mathbf{q} \in \tilde{\eta} \supset \eta^{t_0}(\mathbf{p}_0)$

$$r(\mathbf{p}, \mathbf{q}) = \text{Cov}(X(\mathbf{p}), X(\mathbf{q})) = \frac{2\sigma(\mathbf{p})\sigma(\mathbf{q})L(\mathbf{p})L(\mathbf{q})}{L^2(\mathbf{p}) + L^2(\mathbf{q})} e^{-\frac{|\mathbf{p}-\mathbf{q}|^2}{L^2(\mathbf{p})+L^2(\mathbf{q})}}.$$

Consider also the non-negative bdd real-valued fun.

$$S(\boldsymbol{\omega}; \sigma^2(\mathbf{p}), L(\mathbf{p})) = \frac{\sigma^2(\mathbf{p})L^2(\mathbf{p})}{2\pi} e^{-L^2(\mathbf{p})\frac{|\boldsymbol{\omega}|^2}{2}}, \quad \boldsymbol{\omega} = (\omega_1, \omega_2) \in \mathbb{R}^2,$$

which depends on $\mathbf{p}$ and is locally a spectrum, i.e., if $\tilde{\eta} = \eta^{t_0}(\mathbf{p}_0)$, then $S(\boldsymbol{\omega}; \sigma^2(\mathbf{p}), L(\mathbf{p}))$ coincides with the spectrum in homogeneous case.

Velocity field

Let $t, s \in \tilde{\eta} \supset \eta^{t_0}(\mathbf{p}_0)$ and $L(\mathbf{p}), \sigma^2(\mathbf{p})$ depend only on $\mathbf{p}$. Then, for $t_0 = 0$ and $0 \leq t \leq s \leq 1$ the motion of the field is modelled by means of a flow of diffeomorphisms

$$\phi : \mathbb{R}^2 \times [0, 1]^2 \to \mathbb{R}^2$$

$\phi(\mathbf{p}, 0, 1) = \phi(\mathbf{p}), \phi(\mathbf{p}, s, s) = \mathbf{p}, \phi(\cdot, t, s) = \phi(\cdot, u, s) \circ \phi(\cdot, t, u)$
that are the solution to the transport equation

$$\phi(\mathbf{p}, t, s) = \mathbf{p} + \int_t^s \mathbf{v}(\phi(\mathbf{p}, t, u), u) du, \quad t < s.$$

$\mathbf{p} = \phi(\mathbf{p}_{ts}, t, s),\ \mathbf{p}_{ts} = \phi^{-1}(\mathbf{p}, t, s),\ t < s$.

Spatio-temporal field I

Let $t_0 = 0,\ t = i\,dt,\ \ dt > 0$ and $X_k^*(\mathbf{p}),\ k = 0, 1, \ldots i,$ be independent, Gaussian with non-stationary spatial covariance and $\sigma^2(\mathbf{p}), L(\mathbf{p})$ taking values at $t_0 = 0$. Then, the field $X(\mathbf{p}, t)$ is defined as

$$X(\mathbf{p}, t) = \rho^i X_0^*(\mathbf{p}_{0t}) + \sum_{k=1}^{i} \rho^{i-k} \sqrt{1-\rho^2} X_k^*(\mathbf{p}_{ut}), \qquad u = k\ dt.$$

Spatio-temporal field II

The field $X(\mathbf{p}, t)$ has covariance

$$\text{Cov}(X(\mathbf{p}, t), X(\mathbf{q}, s)) = e^{-\frac{t+s}{2C}} r(\mathbf{p}_{0t}, \mathbf{q}_{0s}) + \\ + \sum_{k=1}^{i} e^{-\frac{t+s-2u}{2C}} \left(1 - e^{-\frac{dt}{C}}\right) r(\mathbf{p}_{ut}, \mathbf{q}_{us}),$$

for $t = i\,dt, s = j\,dt, u = k\,dt$ and $i \leq k \leq j$. When $\mathbf{v}$ is not constant, covariance depends on dt. Letting $dt \to 0$, and $1 - \rho^2 \approx dt/C$, we have

$$\lim_{dt \to 0} \text{Cov}(X(\mathbf{p}, t), X(\mathbf{q}, s)) = e^{-\frac{t+s}{2C}} r(\mathbf{p}_{0t}, \mathbf{q}_{0s}) + \\ + \frac{1}{C} \int_t^s e^{-\frac{t+s-2u}{2C}} r(\mathbf{p}_{ut}, \mathbf{q}_{us})\, du.$$

Modelling significant wave height in the North Atlantic I

For fixed t_0, $\epsilon(\mathbf{p}, t_0) = Y(\mathbf{p}, t_0) - m(\mathbf{p}, t_0)$ is homogeneous over squares $\eta^{t_0}(\mathbf{p}_0)$ with side $\approx 4^\circ$. Moreover, it has been shown that

$$\epsilon(\mathbf{p}, t_0) = \sqrt{p} X_l(\mathbf{p}, t_0) + \sqrt{1-p} X_s(\mathbf{p}, t_0) + X_e(\mathbf{p}, t_0)$$

where $p \in [0, 1]$ mixing parameter and component fields are independent, homogeneous Gaussian fields.
For constant velocity field $\mathbf{v}$ for $t = i\, dt$ and $t_0 = 0$:

$$\epsilon(\mathbf{p}, t) = \rho^i \left[\sqrt{p} X^*_{l,0}(\mathbf{p} - \mathbf{v}t) + \sqrt{1-p} X^*_{s,0}(\mathbf{p}) + X^*_{e,0}(\mathbf{p}) \right] + \sqrt{1-\rho^2} \cdot \sum_{k=1}^{i} \rho^{i-k} \left[\sqrt{p} X^*_{l,k}(\mathbf{p} - \mathbf{v}(t-u)) + \sqrt{1-p} X^*_{s,k}(\mathbf{p}) + X^*_{e,k}(\mathbf{p}) \right],$$

where $u = k\, dt$.

Modelling significant wave height in the North Atlantic II

A natural extension of $\epsilon(\mathbf{p}, t),\ t_0 = 0,\ t = i\,dt,\ i > 0$ is

$$\epsilon(\mathbf{p}, t) = \rho^i \left[\sqrt{p} X^*_{l,0}(\mathbf{p}_{0t}) + \sqrt{1-p} X^*_{s,0}(\mathbf{p}) + X^*_{e,0}(\mathbf{p})\right]$$

$$+ \sum_{k=1}^{i} \rho^{i-k} \sqrt{1-\rho^2} \left[\sqrt{p} X^*_{l,k}(\mathbf{p}_{ut}) + \sqrt{1-p} X^*_{s,k}(\mathbf{p}) + X^*_{e,k}(\mathbf{p})\right],$$

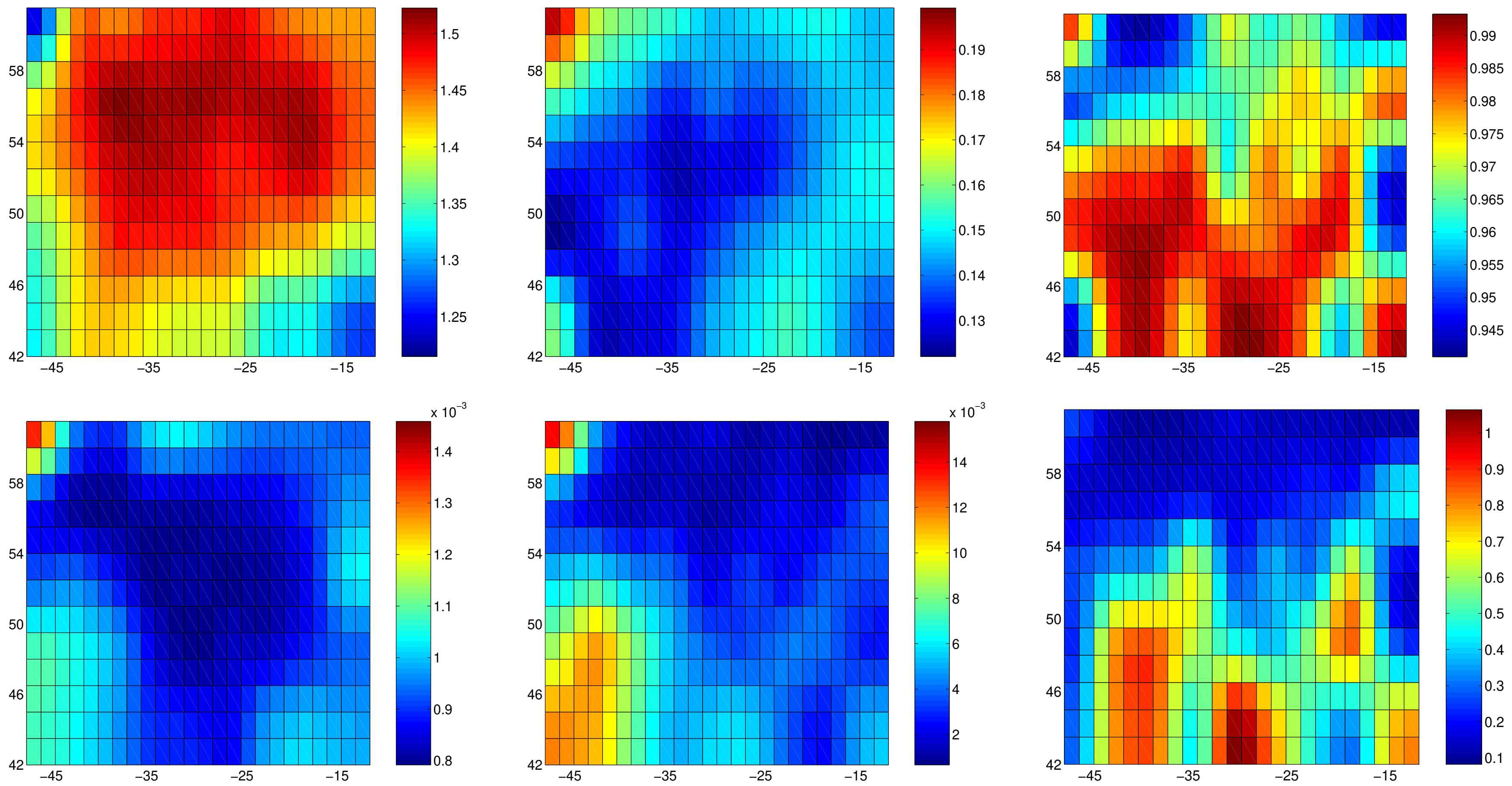
with $u = k\,dt$ and covariance

$$\mathrm{Cov}(\epsilon(\mathbf{p}, t), \epsilon(\mathbf{p}, s)) = \rho^{i+j} \left[p r_l(\mathbf{p}_{0t}, \mathbf{q}_{0t}) + (1-p) r_s(\mathbf{p}, \mathbf{q}) + r_e(\mathbf{p}, \mathbf{q})\right] +$$

$$+ \sum_{k=1}^{i} \rho^{i+j-2k} (1-\rho^2) \left[p r_l(\mathbf{p}_{ut}, \mathbf{q}_{us}) + (1-p) r_s(\mathbf{p}, \mathbf{q}) + r_e(\mathbf{p}, \mathbf{q})\right] =$$

$$= p\mathrm{Cov}(X_l(\mathbf{p}, t), X(\mathbf{q}, s)) + \rho^{j-i} \left[(1-p) r_s(\mathbf{p}, \mathbf{q}) + r_e(\mathbf{p}, \mathbf{q})\right].$$

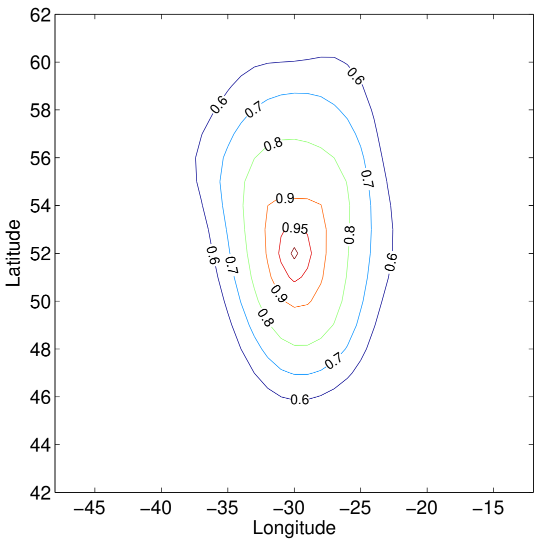
Parameters

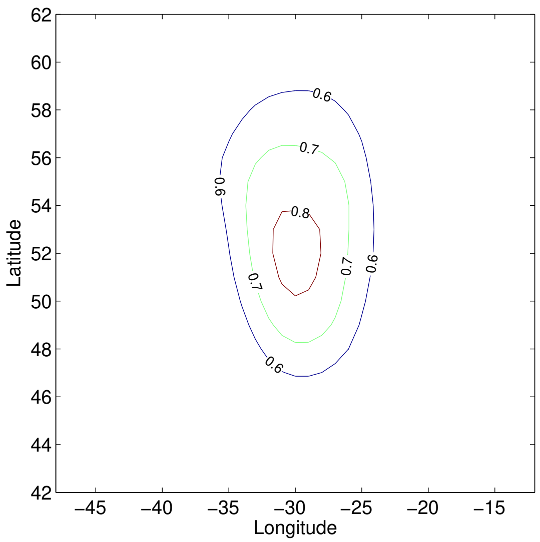


$$m(\mathbf{p},t);\ \sigma^2(\mathbf{p},t);\ p(\mathbf{p},t);\ \sigma_e^2(\mathbf{p},t);\ L_l(\mathbf{p},t);\ L_s(\mathbf{p},t).$$

Covariance function

- C_l, T_l are estimated from buoy, $C_l = 35$ hs, $C_s = 2$min and $C_e = 7$ sec
- Distance between two points cannot exceed 3 degrees $\Rightarrow \forall\; 0 \leq u \leq t,\; \mathbf{p}_{ut} = \mathbf{p} - \mathbf{v}(\mathbf{p})(t-u)$



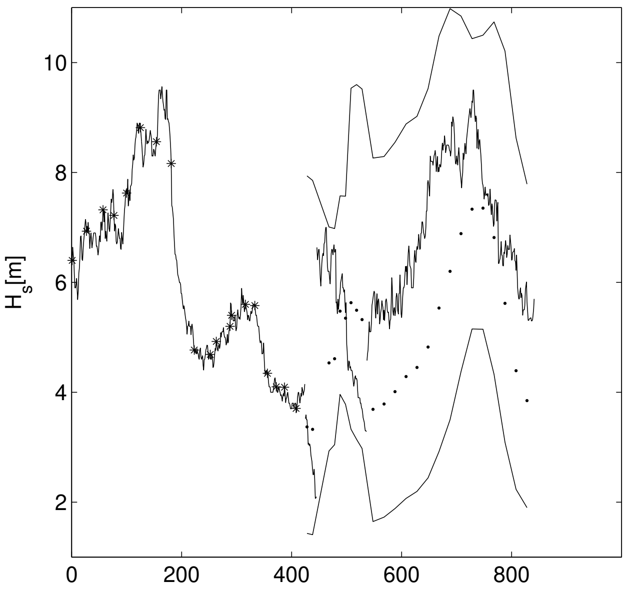


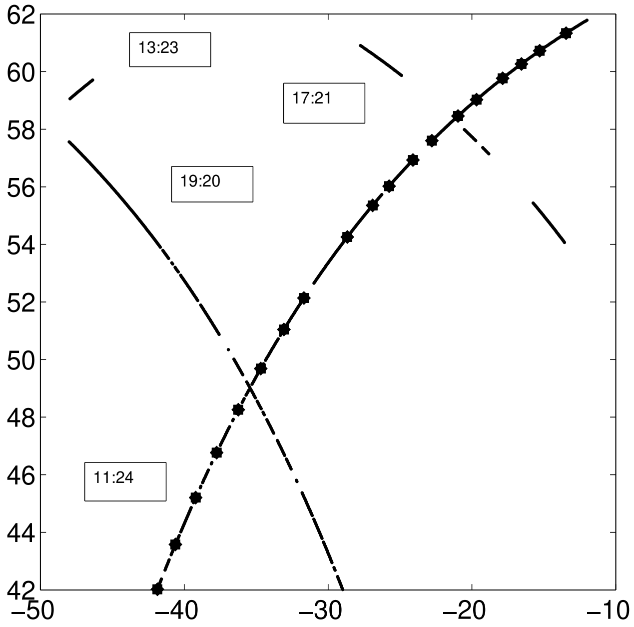
Surface reconstruction

Suppose we want to reconstruct H_s at location $(\mathbf{p}_0, t_0)$. We should check if there are any satellite observations in $\eta(\mathbf{p}_0, t_0)$

- If no, $Y(\mathbf{p}_0, t_0) = m(\mathbf{p}_0, t_0)$
- If yes, the non-homogeneous field X can be predicted by means of the conditional expectation of $X(\mathbf{p}_0, t_0)$ given the observed values of H_s

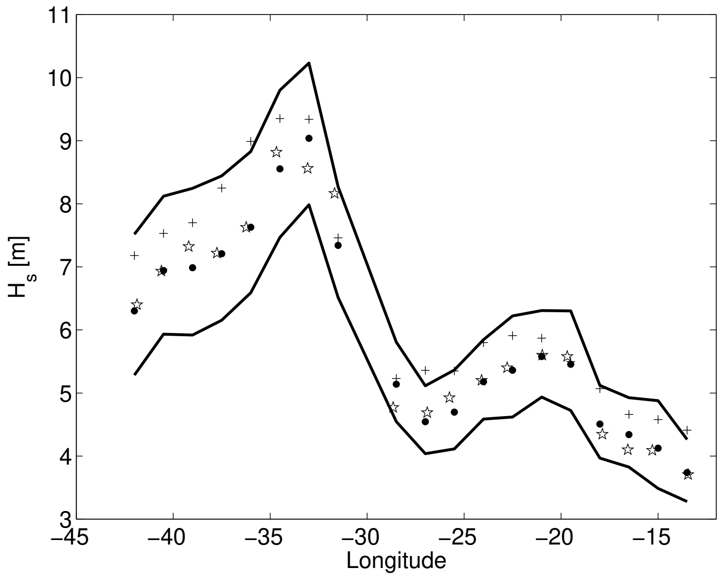
Surface reconstruction

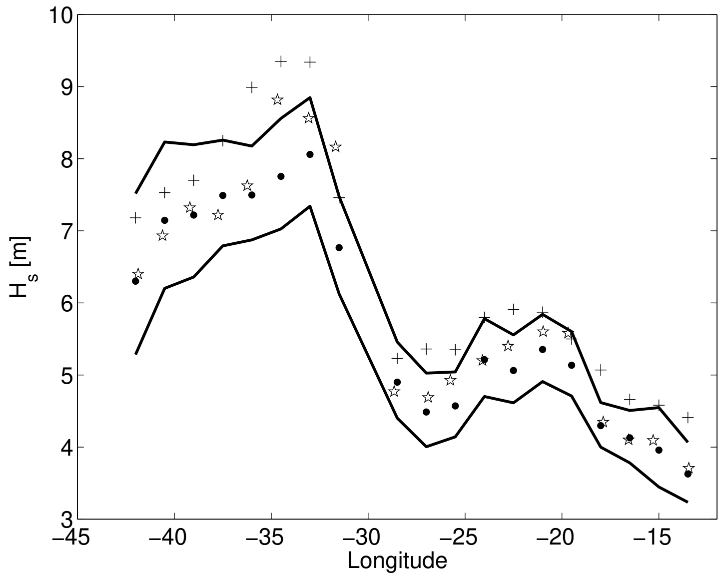




Satellite measurements are the irregular line, stars for the 19 values used to predict the future observations of H_s, the 24 predicted values are black dots and the wide lines are the 99% prediction bands.

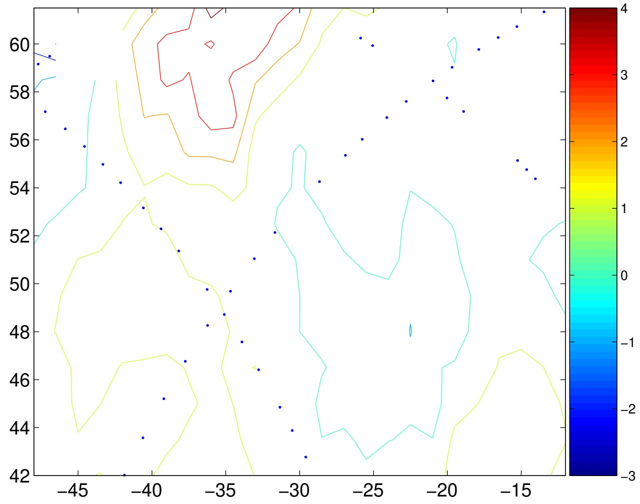
Surface reconstruction





(*Left*) - Values of ERA-40 H_s at 12:00 ('+'); satellite measurements of H_s about 40 minutes earlier (stars); predictions $\hat{H}_s$ at 12:00 at the locations of C-ERA-40 using the 19 satellite observations (dots), black solid lines the 99% prediction bands. (*Right*) - Same as in the (*Left*) plot with $\hat{H}_s$ computed using 43 satellite measurements.

Surface reconstruction



Difference between the C-ERA-40 data and predictions of H_s in the region at 12:00 based on 43 satellite measurements at locations marked by dots.