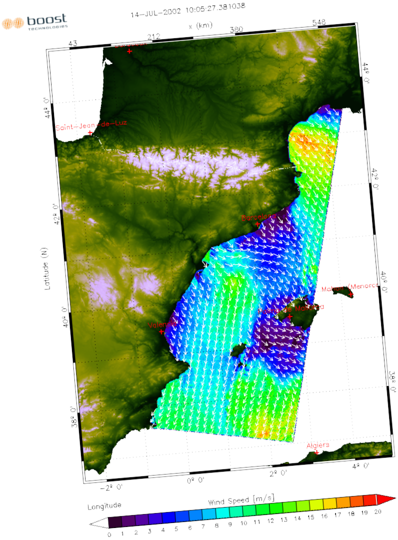
Statistical Emulation of High-Resolution SAR Wind Fields from Low-Resolution Model Predictions

Liyun He Guelton

IOVWST Meeting, June 2014

Ifremer: Bertrand Chapron, Jean Tournadre

Télécom Bretagne: Ronan Fablet



Source: *CLS (ex. Boost Technologies)*

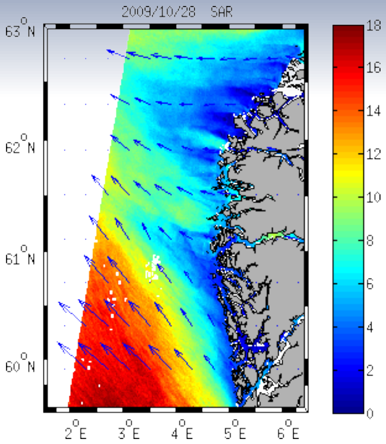
High-Resolution Data Application

- Describe and analyze the ocean dynamics in detail, particularly in coastal area
- Extend and improve numerical prediction models
- Aid energy production, track oil spill disasters
- Help the assessment of risks relevant to marine engineering, environment pollution
- ...

Motivation (1)

Satellite SAR (Synthetic Aperture Radar) Systems

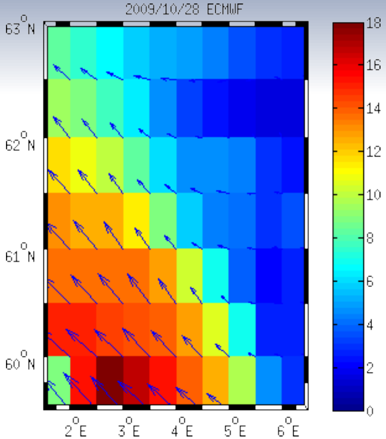
- ☺ High spatial resolution, $< 0.01°$, $\approx 1\,\mathrm{km}$
- ☹ Irregular sampling for a given region
- ☹ Low temporal resolution, every 7-to-10 days for temperate zones



Motivation (2)

Numerical Model Predictions

- ☺ High temporal resolution, every $3\,\mathrm{h}$ or $6\,\mathrm{h}$
- ☺ Global coverage
- ☹ Low spatial resolution, $0.5°$ for example



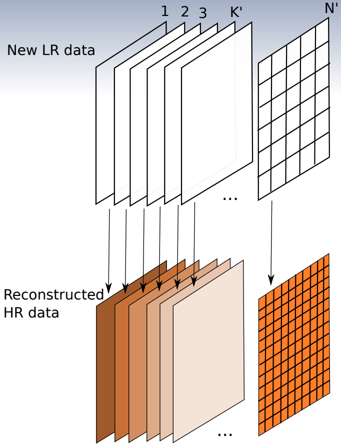
Motivation (3)

Use statistical emulation model

- to learn low-high resolution relationship

For new input LR data

- Derive HR (High Resolution) information from LR (Low Resolution) information
- Reconstruct HR wind fields for anywhere and at anytime



Problem Formulation

Regression Problem

- $y = f(x)$
- x and y are two-dimensional vector fields parameterized according to the zonal and meridional wind components

Problem Formulation

Regression Problem

- $y = f(x)$
- x and y are two-dimensional vector fields parameterized according to the zonal and meridional wind components

Learning Scheme

- Goal: learn regression function f
- From a training set $\{(x_k, y_k)\}$, $k \in (1, n)$
- Find an optimal f^*: minimize the regression errors $\sum_{k=1}^{n}(y_k - f^*(x_k))$ or $\sum_{k=1}^{n}(y_k - f^*(x_k))^2$

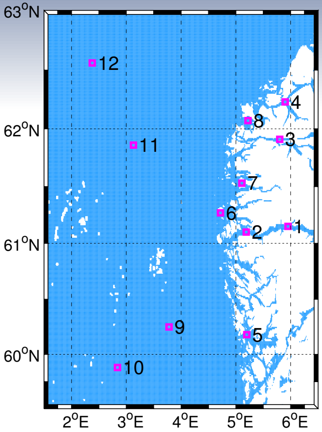
Data Set and Study area

Data set

- Select the closest ECMWF data to SAR data acquisition ; very different pairs are rejected
- 758 pairs for the period 2005–2010

Study Area

- Southwestern coastal sea of Bergen, Norway
- Many islands, mountains, and fjords in coastal area



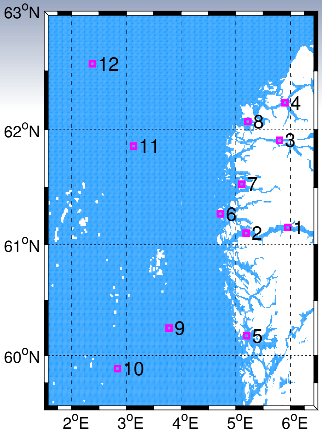
Data Set and Study area

Data set

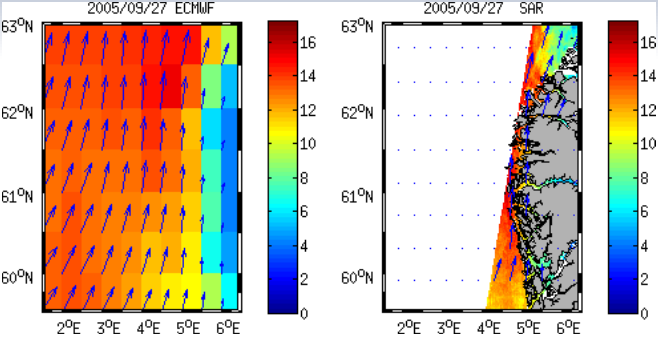
- Select the closest ECMWF data to SAR data acquisition ; very different pairs are rejected
- 758 pairs for the period 2005–2010

Study Area

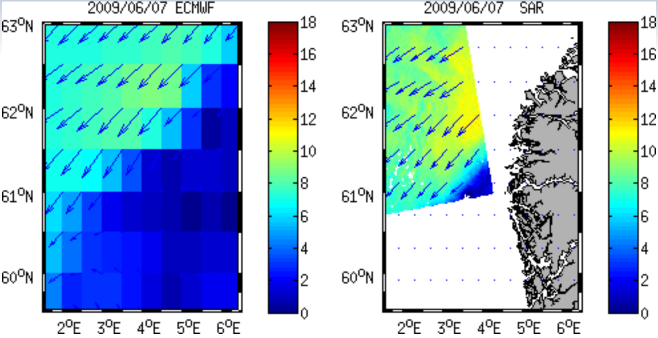
- Southwestern coastal sea of Bergen, Norway
- Many islands, mountains, and fjords in coastal area ⇒ Remarkable local effects: coastal jets, wind shadows, land-sea breezes. . .



Data Set Examples (1)

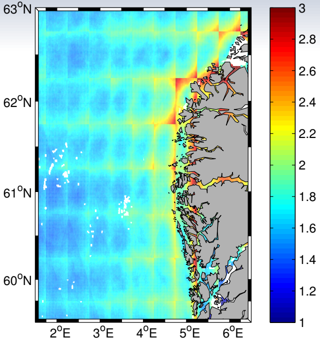


Data Set Examples (2)

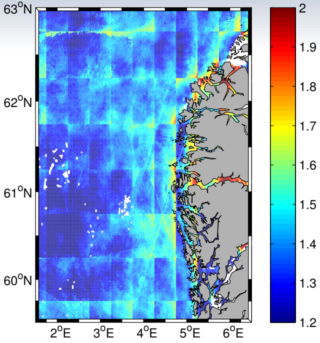


Physical Constraint Analyses (1)

Global Analyses



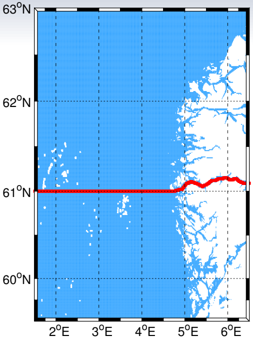
Mean of the difference between ECMWF and SAR



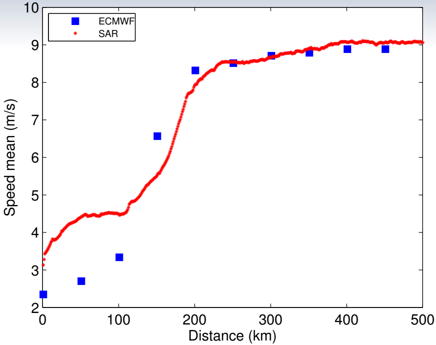
Standard deviation of the difference ECMWF-SAR

Physical Constraint Analyses (2)

Transect perpendicular to the coast



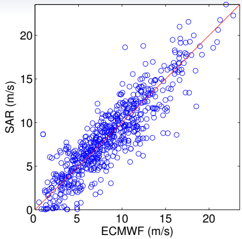
Transect line



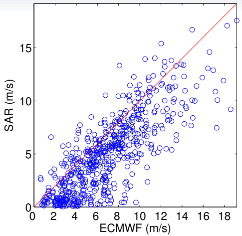
ECMWF and SAR mean

Physical Constraint Analyses (3)

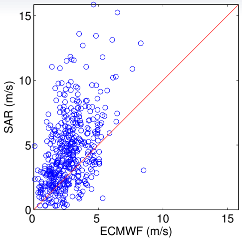
Scatterplots of ECMWF-SAR for: offshore, coast, fjord feature grid point



Offshore feature point 9
(N62.25°, E3.78°)



Coastal feature point 8
(N62.07°, E5.22°)

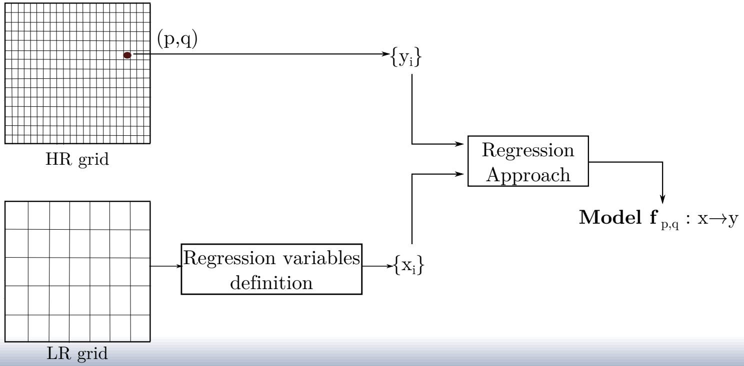


Fjord feature point 1
(N61.15°, E5.95°)

Non-linear LR-HR relationship in coastal and fjord area, compared to offshore area

Proposed Regression Model

- Point-specific model: a transfer function for each HR grid point
- Regression approach choice
- Regression variables (predictors) definition



Regression Technique Choices

- Analog methods – standardized responses of similar cases:

$$f(x) = \sum_{s=1}^{n} w_s \, \mathrm{g}(x, x_s) \tag{1}$$

- Multiple Linear Regression (MLR) – linear relation between inputs and outputs:

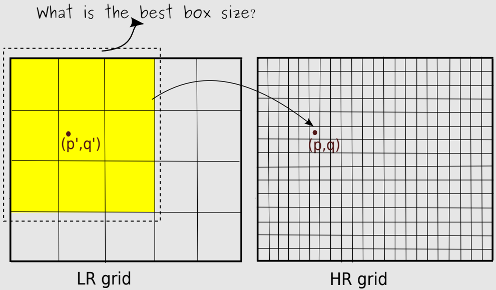
$$f(x) = \omega^t x + b \tag{2}$$

- Non-Linear SVR (Support Vector Regression) – optimal non-linear kernel-based model, linear regression model in a space defined by a non-linear function mapping:

$$f(x) = \omega^t \Phi(x) + b \tag{3}$$

Regression Variable Definition (1)

- Global information: the first *m* Principle Components (PCs) in the Empirical Orthogonal Functions (EOFs) space
- Local information: exploits the LR wind information within a local neighborhood



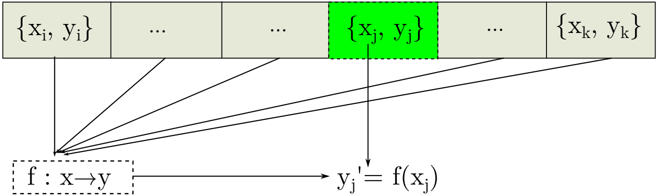
Regression Variable Defintion (2)

- Entropy-based information selection step:
  - Conditional entropy $H(y|x)$ is a measure of amount of uncertainty remaining about y after x is known
  - For a given HR grid point p, we select the LR grid points with the lowest conditional entropy values for HR field at this point:

$$H(y_p|x_q) = -\sum_{j=1}^{m}\sum_{i=1}^{n} P(y_p = y_j, x_q = x_i) \log P(y_p = y_j|x_q = x_i) \quad (4)$$

Evaluation

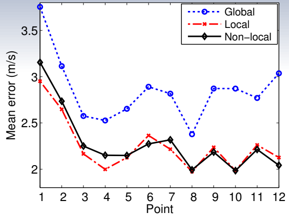
- Qualitative and quantitative evaluation
- Cross validation
  - 95% randomly sampled for training
  - 5% for reconstruction



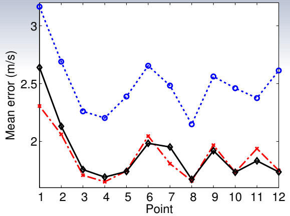
- Regression performance for 12 HR grid points

Regression Variable Influence

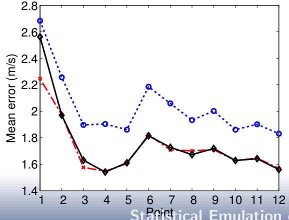
Information type: global, local and entropy-based information



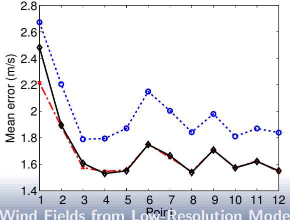
Nearest Neighbor



K-Nearest Neighbor



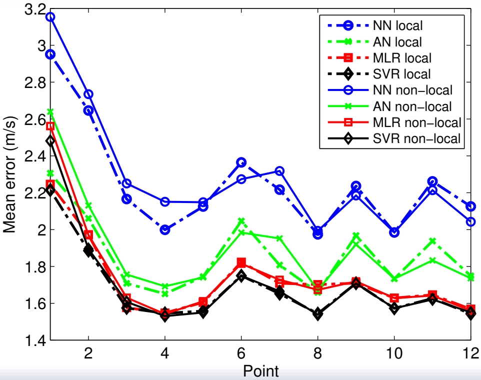
MLR



SVR

Regression Approach Influence

Compared approaches: Nearest Neighbor (NN), Weighted Average Analog (AN), Multiple Linear Regression (MLR), Support Vector Regression (SVR)

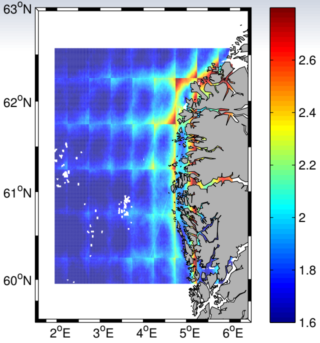


Optimal Regression Model

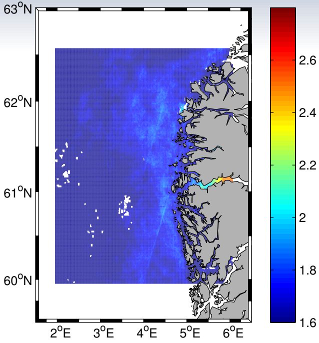
- SVR
- Local or non-local entropy-based variables' selection

Statistical Properties (1)

Global Analyses



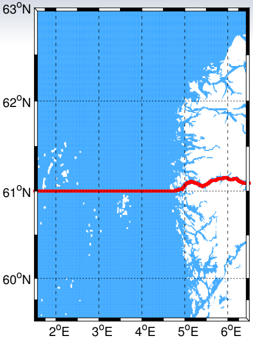
Mean of the difference between ECMWF and SAR data



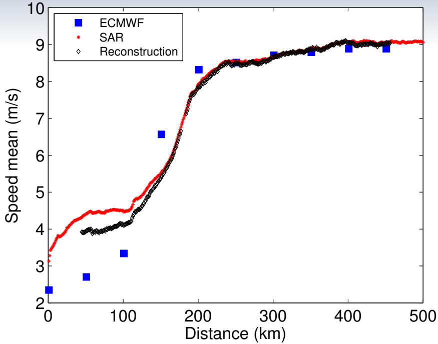
Mean of the difference between Reconstruction and SAR data

Statistical Properties (2)

Transect perpendicular to the coast



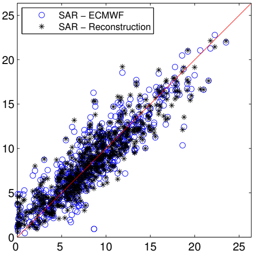
Transect line



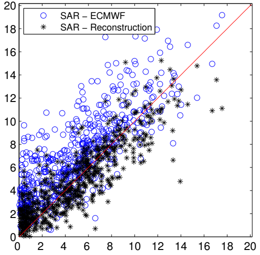
ECMWF, SAR and Reconstruction speed mean

Statistical Properties (3)

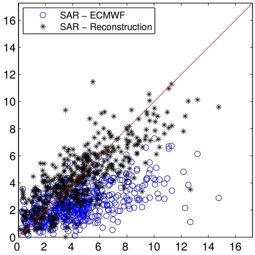
Scatterplots of SAR-ECMWF and of SAR-Reconstruction for: offshore, coast, fjord feature grid point



Offshore feature point 9
(N62.25°, E3.78°)

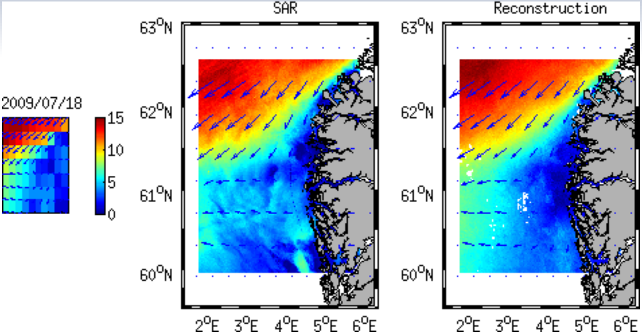


Coastal feature point 8
(N62.07°, E5.22°)

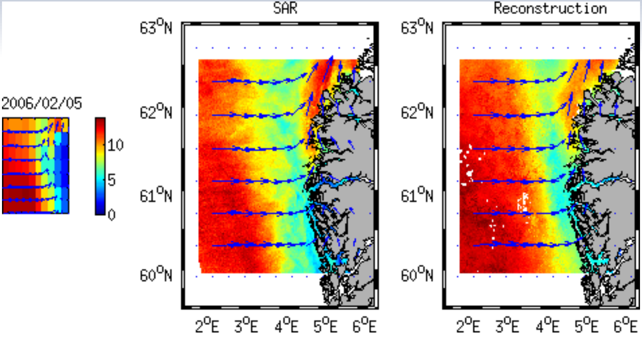


Fjord feature point 3
(N61.91°, E5.80°)

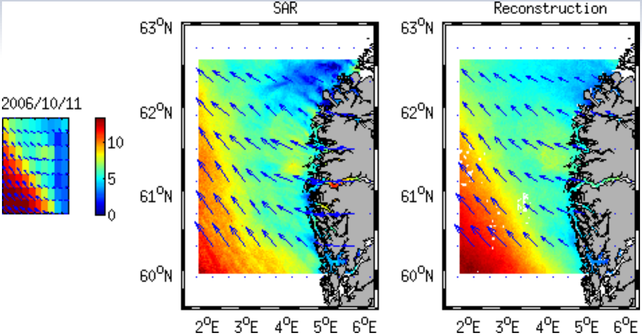
Reconstruction Example (1)



Reconstruction Example (2)



Reconstruction Example (3)



Thanks!