Statistical High Resolution Wind Field Emulation

Liyun Guelton

Ifremer / Télécom Bretagne

July 2013
Thesis Goal

Develop original and efficient solutions for the interpretation and spatial interpolation of high resolution oceanic dynamics,

Using a set of history observations of ocean surface, if possible, to take into account dynamics constraints.
Available information

- LR (*Low Resolution*) at time \( t \)
- Set of pairs of high-low resolution data
Available information

✓ LR (Low Resolution) at time $t$
✓ Set of pairs of high-low resolution data

Data set — wind fields

- Low resolution – ECMWF analyse: spatial resolution - 50 km; temporal resolution - 6 h
- High resolution – SAR Envisat: spatial resolution - 1 km; irregular temporal resolution; partial spatial coverage
Available information

✔ LR (Low Resolution) at time $t$
✔ Set of pairs of high-low resolution data

Data set — wind fields

► Low resolution – ECMWF analyse: spatial resolution - 50 km; temporal resolution - 6 h
► High resolution – SAR Envisat: spatial resolution - 1 km; irregular temporal resolution; partial spatial coverage

maximum 3 h time shift $\Rightarrow$ difficult time matching
Context

Study Area

- south-west of Norway
- Longitude 1.5° – 6.5°
- Latitude 59.5° – 63.0°

Area features

- Many fjords
- Many mountains
LR and HR Examples

2009/01/24 - ECMWF

2009/01/24 - SAR

Latitude (°)

Longitude (°)

0 2 4 6 8 10 12

0 2 4 6 8 10 12
Data Set Examples

[Image of data set examples with two graphs showing geographic data over a defined area with color scales and arrow vectors indicating direction and magnitude.]
Feasibility?

Physical link

- LR – general atmospheric circulation
- HR – conditioned by general atmospheric circulation, but resulting from local factors

Data analyse – relationship between HR and LR in data set

- Local link
- Global link
Correlation for an offshore point

Local link
Good correlation in offshore
Correlation for a fjord point

Local link
Not so good onshore and inshore, because of
- non-linearity between LR and HR?
- bad representativeness of LR?
- both of them
Entropy – information given by one point to another one

Global link

- Compute the entropy between one point and all other points around;
- Onshore and inshore point: bad local entropy, good entropy with faraway points

Offshore Point

fjord Point
Physical constraints

Non stationary in the spatial domain?

✓ Average and variance for different points;
✓ Average from the coast to offshore; global and by direction;
✓ PDF for different points: wind rose, histogram;
Histogram of the difference between ECMWF and SAR for each point of each zone: fjord, inshore, offshore.
Difference between SAR and ECMWF depending on the distance to the coast
Wind Rose

ECMWF Point 1

SAR Point 1

ECMWF Point 11

SAR Point 11
Downscaling Approach

Definition
A process that links variables that represent large scale and small scale information

Concept

- **Large Scale** – circulation pattern over a large zone; it slowly variates in the spatial domain (smooth contour of the correlation)
- **Small scale** – local informations in a point; taken from *in situ* measurements or from satellite high-resolution observations
- **Link** – it should be fundamental and physical
Principal Approaches

Dynamic Downscaling

- Local modeling using physics law
- 😊 Realistic modelling of climatic phenomenon
- 😞 Had to apply to a complex process; irrelevant for climatic changes; CPU time consuming

Statistical Downscaling

- Quantitative relation between high and low scale using machine learning on historical data
- 😊 Easy to implement, low cost after the learning phase
- 😞 supposing that the model is stationary; performance limited by data set representiveness
Statistical Downscaling

We are only interested in methods to get the transfer function $f$ to be able to directly predict $Y$

$$Y = f(x, E)$$ (1)

Method’s View

- Linear Regression – linear relation between inputs and response:
  $$f(x) = \omega^t x + b$$ (2)

- Non-Linear Regression – non-linear function transfer:
  $$f(x) = \omega^t \Phi(x) + b$$ (3)
Model’s View

- Global Model: a single transfer function $f$
- Local Model: a transfer function per point

Choice

✗ Global Model, Linear regression: $SVD$ (Singular Value Decomposition)
✗ Global Model, Non-Linear regression: Analog Methods; Classifying Methods
✓ Local Model, Linear regression: $MLR$ (Multiple Linear Regression)
✓ Local Model, Non-Linear regression: $SVR$ (Support Vector Regression)
?
Local Model, Hybrid regression
Methods Comparison

Linear Regression

😊 Reasonable approximate; low CPU usage
😊 Remove all variation types; inadequate for non-linear problems

Non-Linear Regression

😊 Can be fairly generalized; keeps natural properties of the data; naturally includes analogue methods and classifying methods
😊 Iterative optimization ⇒ highly CPU demanding training

Hybrid Methods

😊 Better global performance
😊 Creates spatial discontinuity
Implementation

- Data between 2005 and 2010
- U and V are independently emulated
- Comparison between predictor choice: local (Neighborhood distance = 1); global (all the study area) or local + global
- PCA on LR to get global information: decreases redundant information to capture global situation
- Comparison between methods: MLR; Classification + MLR (proposed); SVR (proposed); Hybrid Method using automatic selection during validation test (proposed)
reduce the noise; reduce the size of predictors
Result of linear regression global and local predictors
Proposed Method

- Local Method: optimize each point
- Use global information to emulate points with low quality LRs
- Classification + linear regression
- Non-Linear Regression: SVR
- Hybrid regression
- Use validation tests to select the type of automatic predictor
Result of classification + linear regression
Result of SVR

Wind behavior in class 1

Wind behavior in class 2

Wind behavior in class 3

Wind behavior in class 4
Hybrid Regression Schema

Learning data for grid point \((i,j)\) in class \(k\)

<table>
<thead>
<tr>
<th>Combination test learning data</th>
<th>Combination test validation data</th>
</tr>
</thead>
<tbody>
<tr>
<td>((X,Y))</td>
<td>(X', Y')</td>
</tr>
</tbody>
</table>

- \(PG 1\) vs \(RM 1\)
- \(PG 2\) vs \(RM 1\)
- \(
\vdots
\)
- \(PG N\) vs \(RM 1\)
- \(PG 1\) vs \(RM M\)
- \(PG 2\) vs \(RM M\)
- \(
\vdots
\)
- \(PG N\) vs \(RM M\)

- model \((1,1)\) \(\rightarrow\) \(MSE_{1,1}\)
- model \((2,1)\) \(\rightarrow\) \(MSE_{2,1}\)
- \(
\vdots
\)
- model \((N,1)\) \(\rightarrow\) \(MSE_{N,1}\)
- model \((1,M)\) \(\rightarrow\) \(MSE_{1,M}\)
- model \((2,M)\) \(\rightarrow\) \(MSE_{2,M}\)
- \(
\vdots
\)
- model \((N,M)\) \(\rightarrow\) \(MSE_{N,M}\)

\(\min\)

Best Combination \(PG s\) vs \(RM t\)

Model \(f_{ij}^k\)

PG: Predictors Group
RM: Regression Method
MSE: Mean Square Error
Hybrid Regression Result

2008/09/26 ECMWF

Latitude

Longitude

0 5 10 15 20 25
Hybrid Regression Result
Hybrid Regression Result
MERCI!
THANK YOU!