

Development, deployment and validation of an oceanographic virtual laboratory based on Grid computing

David Mera Pérez
Brno, Czech Republic, October 21th 2013

Index

1. Context and motivation
2. Main goals
3. Retelab project - Virtual laboratory development
4. Sentinazos project - Virtual laboratory validation
5. On going work



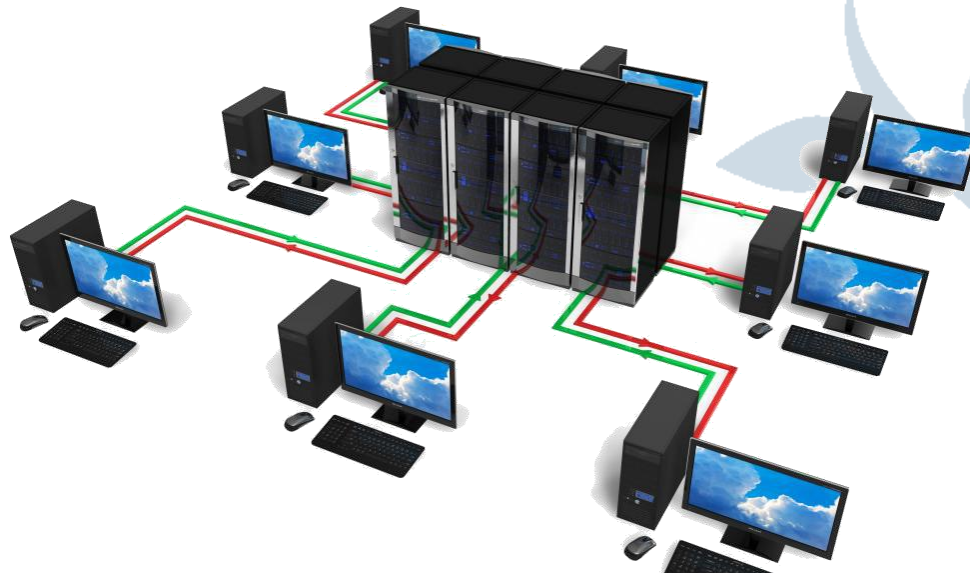
Index

1. **Context and motivation**
2. Main goals
3. Retelab project - Virtual laboratory development
4. Sentinazos project - Virtual laboratory validation
5. On going work



Context and Motivation

- The oceanographic research community has access to huge amount of available datasets
- Researchers must deal with new issues such as how to store, process and make the best of available datasets
- The study of the ocean requires multidisciplinary teams
- Several levels of computer skills



Context and Motivation

- Distributed computing: **Grid computing**
 - Resource sharing via Internet
 - Low cost
 - Security
 - Open Standards
 - Virtual Organizations management



Index

1. Context and motivation
- 2. Main goals**
3. Retelab project - Virtual laboratory development
4. Sentinazos project - Virtual laboratory validation
5. On going work



Main goals

- To develop a user-friendly distributed computational environment based on Grid computing.
- To develop an oceanographic application to test the Grid environment.
 - An oil spill automatic detection system based on the analysis of satellite Synthetic Aperture Radar imaging.



RETELAB



Index

1. Context and motivation
2. Main goals
- 3. Retelab project - Virtual laboratory development**
4. Sentinazos project - Virtual laboratory validation
5. On going work



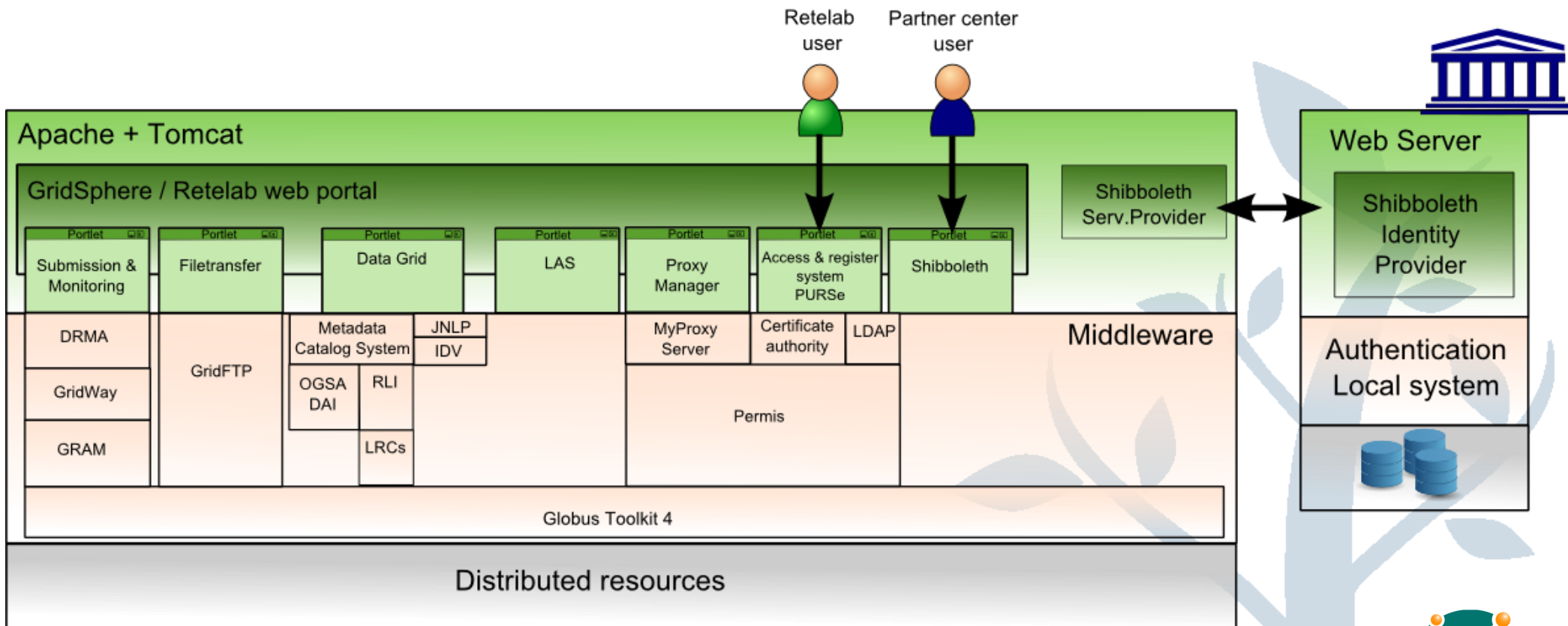
Retelab

Basis

- Easy and accessible tool
- Computer skills should be minimum or even unnecessary
- Open source
- Security
- Distributed storage and computational capacity



Retelab Architecture



Retelab

User access and registration system

Traditional Grid systems

- Command line interface
- Digital certificates managed by users
- Advanced computer skills are required
- System user registration
- Complex user account management

vs

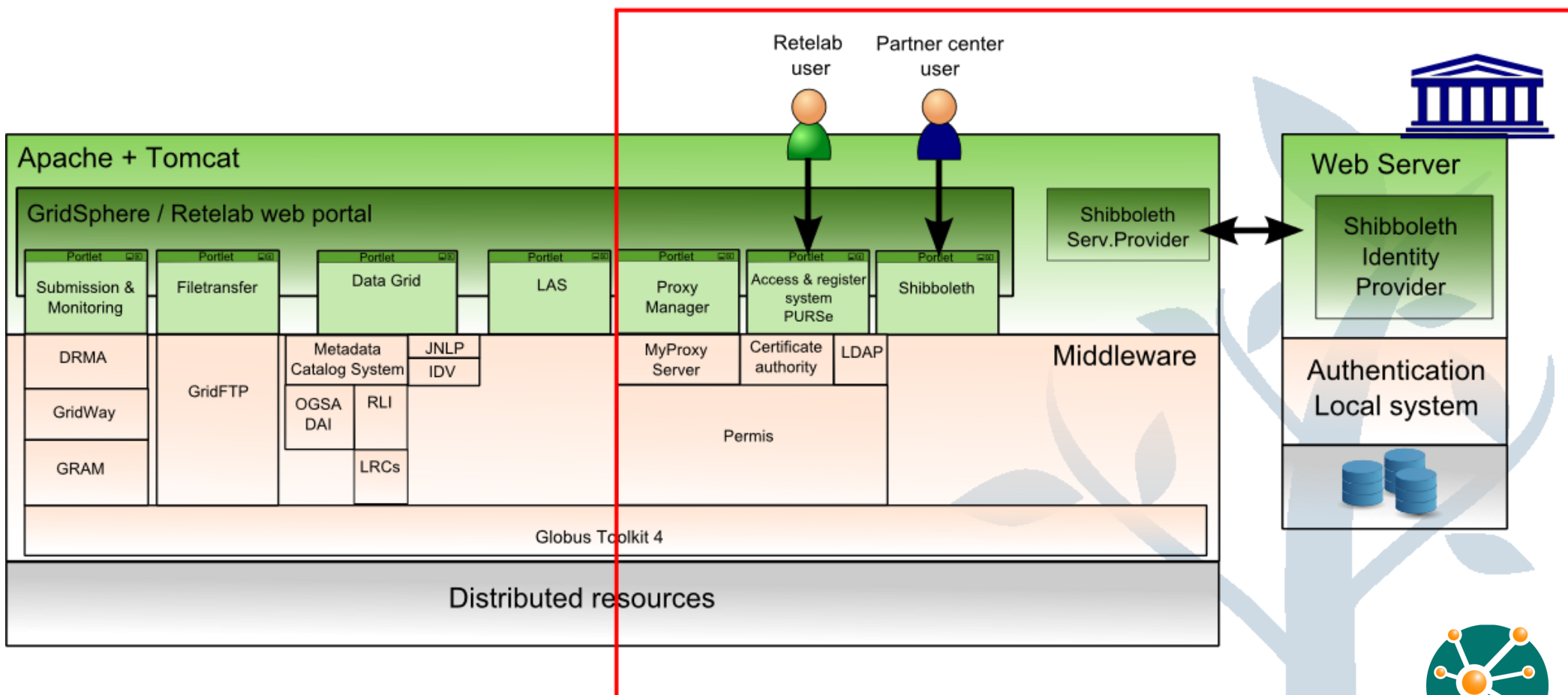
RETELAB

- Web interface based on portlets
- Credential managed by the system
- Minimum computer skill are required
- Single Sign-On registration system
- Role base access system

Retelab

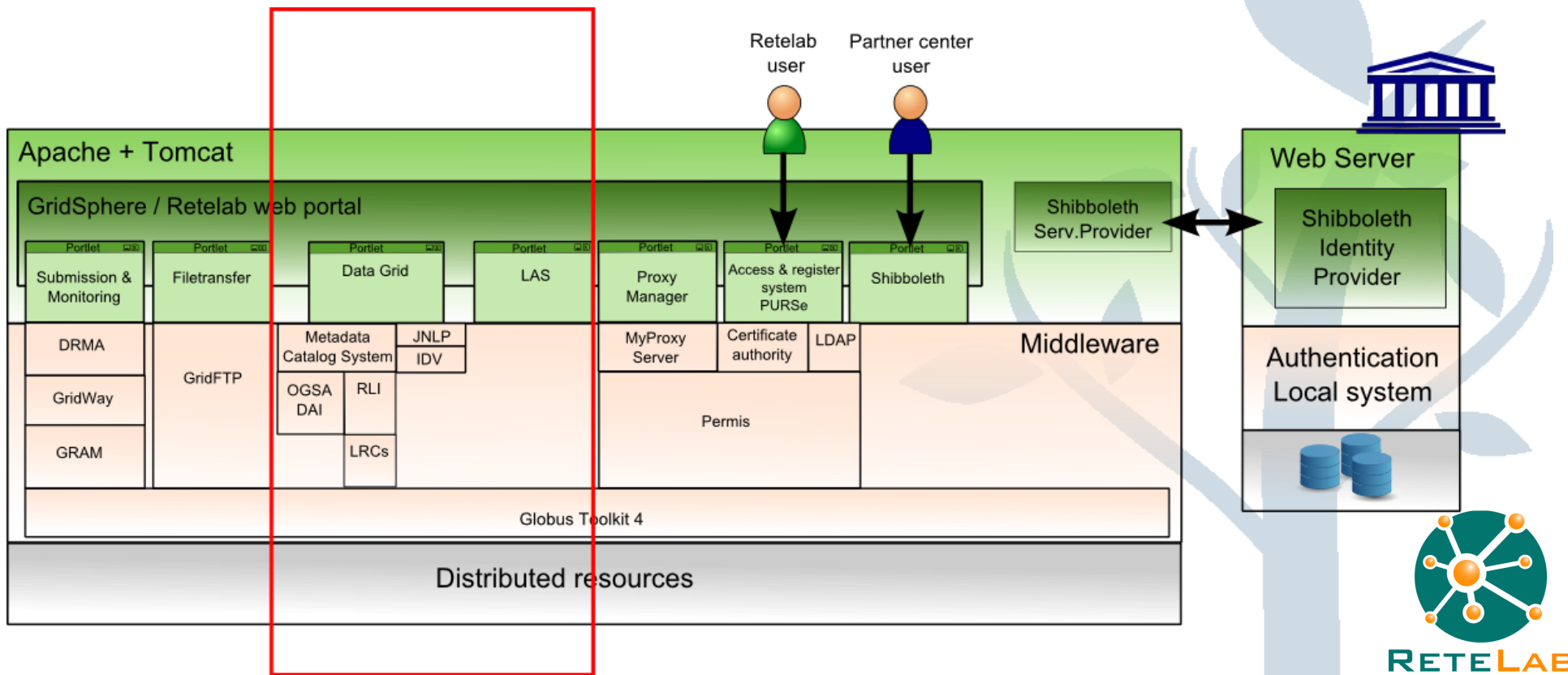
User access and registration system

- Arquitectura



Retelab Distributed storage system

- Web interface integration
- Based on Metadata - ISO 19115, Geographic Information
- Integration of visualization tools



Retelab

Distributed storage system

- Visualization tools

Integrated Data Viewer

Live Access Server

Retelab

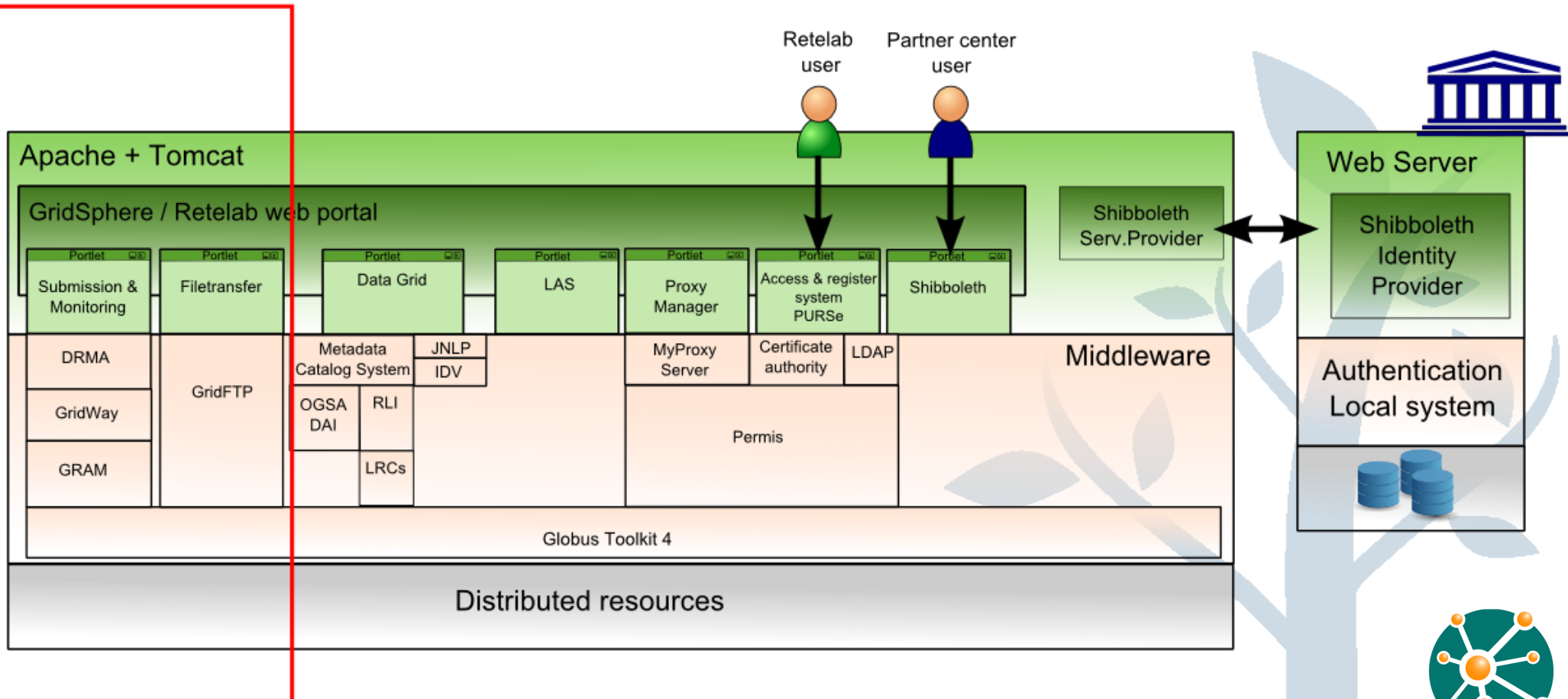
Job submission and monitoring system

- Traditional Grid systems
 - User interaction was required
 - Complex submission systems
- **RETELAB**
 - Grid metascheduler
 - Web integration
 - To make decisions on behalf of users
 - To facilitate the optimal utilizations of available resources
 - It undertakes the tasks for resource discovery, job scheduling, executing, monitoring and output retrieval.



Retelab

Job submission and monitoring system



Retelab Use case



Produccion primaria

IDL - PRIMARY PRODUCTION

Title:

Main:

Input files:

Result files:

Remote:

Local:

Enviar

Virtual Data Base

Region = ▾ Canary

Sensor = ▾

Parameter = ▾

Search

↑ :: /home/retelab/david_mera

Select input files

- loadct.pro
- pp_out_nc.pro
- aware.pro
- fotoper.pro

Retelab Use case



Produccion primaria

IDL - PRIMARY PRODUCTION

Title: `virtual9`

Main: `aware`

Input files:
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/AV_SST_2006314_6_600.N7.newmedvirado`
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/SW_CHL_2006314_6_600.14.newmedvirado`
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/S20063132006320.L3m_8D_PAR_9 file:///home/r`

Result files:
Remote: `results.tar.gz`
Local: `results.tar.gz`

Virtual Data Base

- AV_SST_2006314_6_600.N7.newmedvirado [+]
- SW_CHL_2006314_6_600.14.newmedvirado [+]
- S20063132006320.L3m_8D_PAR_9_2 [+]
- sw_ppo_2006314_6_600.14.newmedvirado.nc [+]

Selecc
Show Search

/home/retelab/david_mera

Select_input_files

- 1243337736781
- 1245862964637
- loadet.no

Enviar

Retelab Use case



Produccion primaria

IDL - PRIMARY PRODUCTION

Title: `virtual9`

Main: `aware`

Input files:
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/AV_SST_2006314_6_600.N7.newmedvirado`
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/SW_CHL_2006314_6_600.14.newmedvirado`
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/S20063132006320.L3m_8D_PAR_9 file:///home/r`

Result files:
Remote: `results.tar.gz`
Local: `results.tar.gz`

Virtual Data Base

- AV_SST_2006314_6_600.N7.newmedvirado [+]
- SW_CHL_2006314_6_600.14.newmedvirado [+]
- S20063132006320.L3m_8D_PAR_9_2 [+]
- sw_ppo_2006314_6_600.14.newmedvirado.nc [+]

Selecc
Show Search

/home/retelab/david_mera

Select_input_files

- 1243337736781
- 1245862964637
- loadet.nro

Enviar

Retelab Use case



Produccion primaria

IDL - PRIMARY PRODUCTION

Title: `virtual9`

Main: `aware`

Input files:

`gsiftp://ui.retelab.cesga.es/tmp/david_mera/AV_SST_2006314_6_600.N7.newmedvirado`
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/SW_CHL_2006314_6_600.14.newmedvirado`
`gsiftp://ui.retelab.cesga.es/tmp/david_mera/S20063132006320.L3m_8D_PAR_9 file:///home/r`

Result files:

Remote: `results.tar.gz`

Local: `results.tar.gz`

Virtual Data Base

- AV_SST_2006314_6_600.N7.newmedvirado [+]
- SW_CHL_2006314_6_600.14.newmedvirado [+]
- S20063132006320.L3m_8D_PAR_9_2 [+]
- sw_ppo_2006314_6_600.14.newmedvirado.nc [+]

Selecc

Show Search

🏠 `/home/retelab/david_mera`

Select_input_files

- 📁 1243337736781
- 📁 1245862964637
- 📄 loadet.pro

Enviar

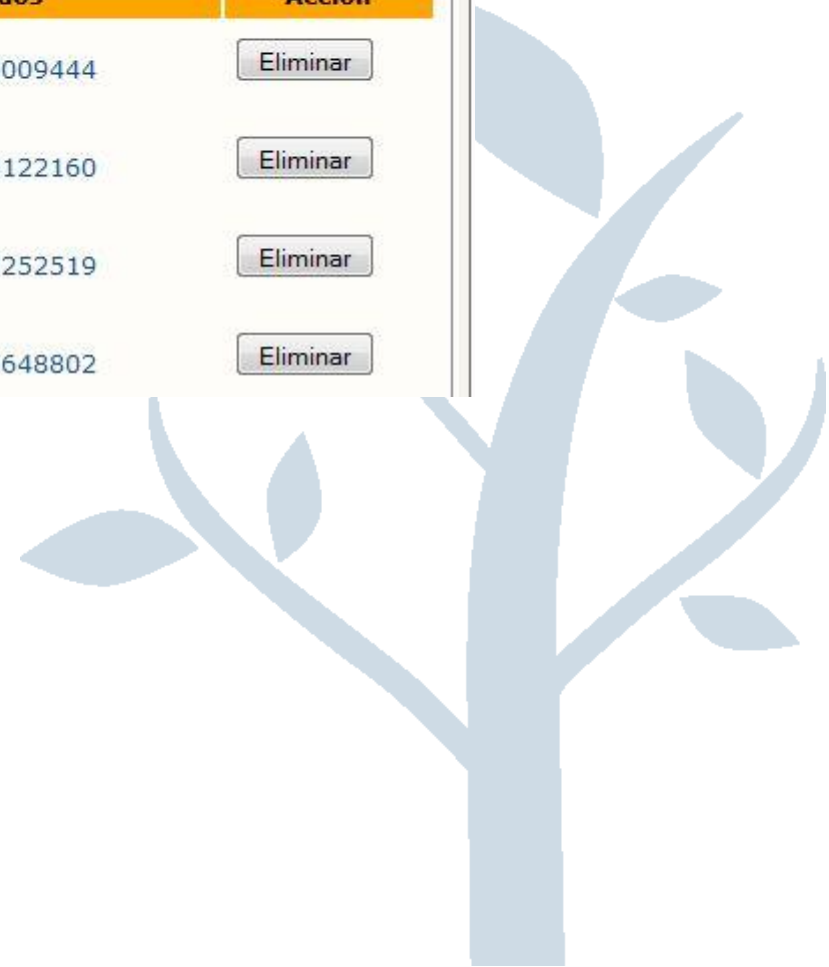
Retelab

Use case



The screenshot shows a web application window titled "GW monitor". Inside the window, there is a section titled "Monitor de trabajos". Below this title is an "Actualizar" button. The main content is a table with the following columns: "GWID", "Trabajo", "Estado", "DN proxy", "Resultados", and "Accion". The table contains four rows of data, all with the state "Finalizado". Each row has an "Eliminar" button in the "Accion" column.

GWID	Trabajo	Estado	DN proxy	Resultados	Accion
5	testtask	Finalizado	testuser9	david_mera/1241005009444	Eliminar
6	testtask	Finalizado	testuser9	david_mera/1241005122160	Eliminar
7	testtask	Finalizado	testuser9	david_mera/1241007252519	Eliminar
8	testtask	Finalizado	testuser9	david_mera/1241007648802	Eliminar



Retelab Use case

GW monitor

Monitor c ?

Actualizar

GWID

5 t

6 t

7 t











8 t

return

Username: david_mera

Id Job: 96

Directory: /home/retelab/david_mera/1245261052597


Type	Name	Size	Last Modification	Actions
	virtual3.jt	1098	20090617175054	 add DataGrid
	idl.in	5	20090617175052	 add DataGrid
	stderr.96.txt	657	20090617175200	 add DataGrid
	stdout.96.txt	759	20090617175159	 add DataGrid
	sw_ppo_2006314_6_.600.14.newmedvirado.nc	4771292	20090617175155	 add DataGrid

Atributos

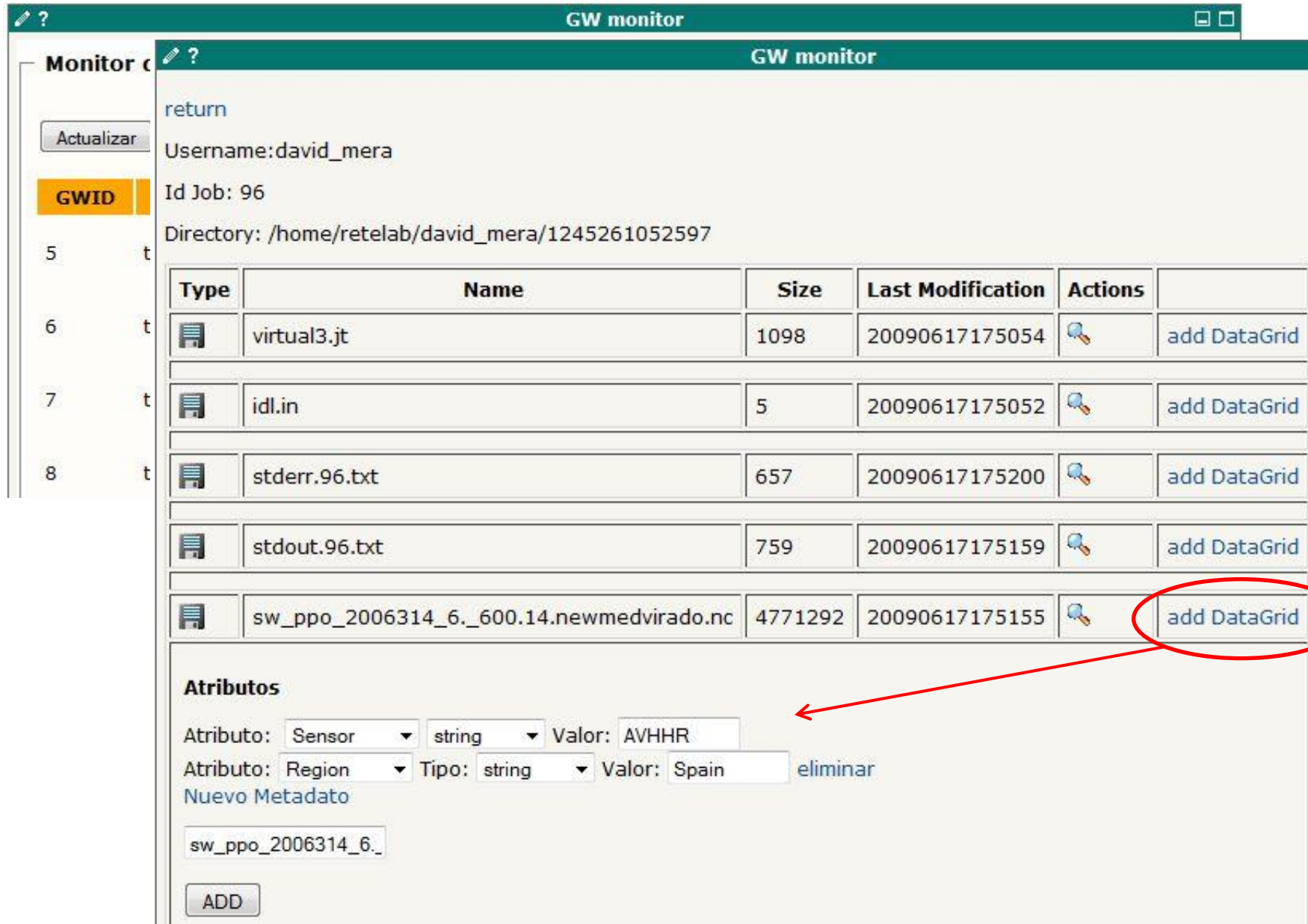
Atributo: Tipo: Valor:

Atributo: Tipo: Valor: [eliminar](#)

[Nuevo Metadato](#)



Retelab Use case



The screenshot displays the 'GW monitor' application interface. On the left, there is a sidebar with a 'Monitor' section containing an 'Actualizar' button and a 'GWID' button. The main content area shows a user profile for 'david_mera' with job ID 96 and a directory path. Below this is a table listing files with columns for Type, Name, Size, Last Modification, and Actions. The 'sw_ppo_2006314_6_600.14.newmedvirado.nc' file is highlighted with a red circle, and a red arrow points from this circle to the 'Atributos' section below. The 'Atributos' section includes fields for 'Sensor' (string, AVHHR) and 'Region' (string, Spain), along with an 'eliminar' button and a 'Nuevo Metadato' section with an 'ADD' button.

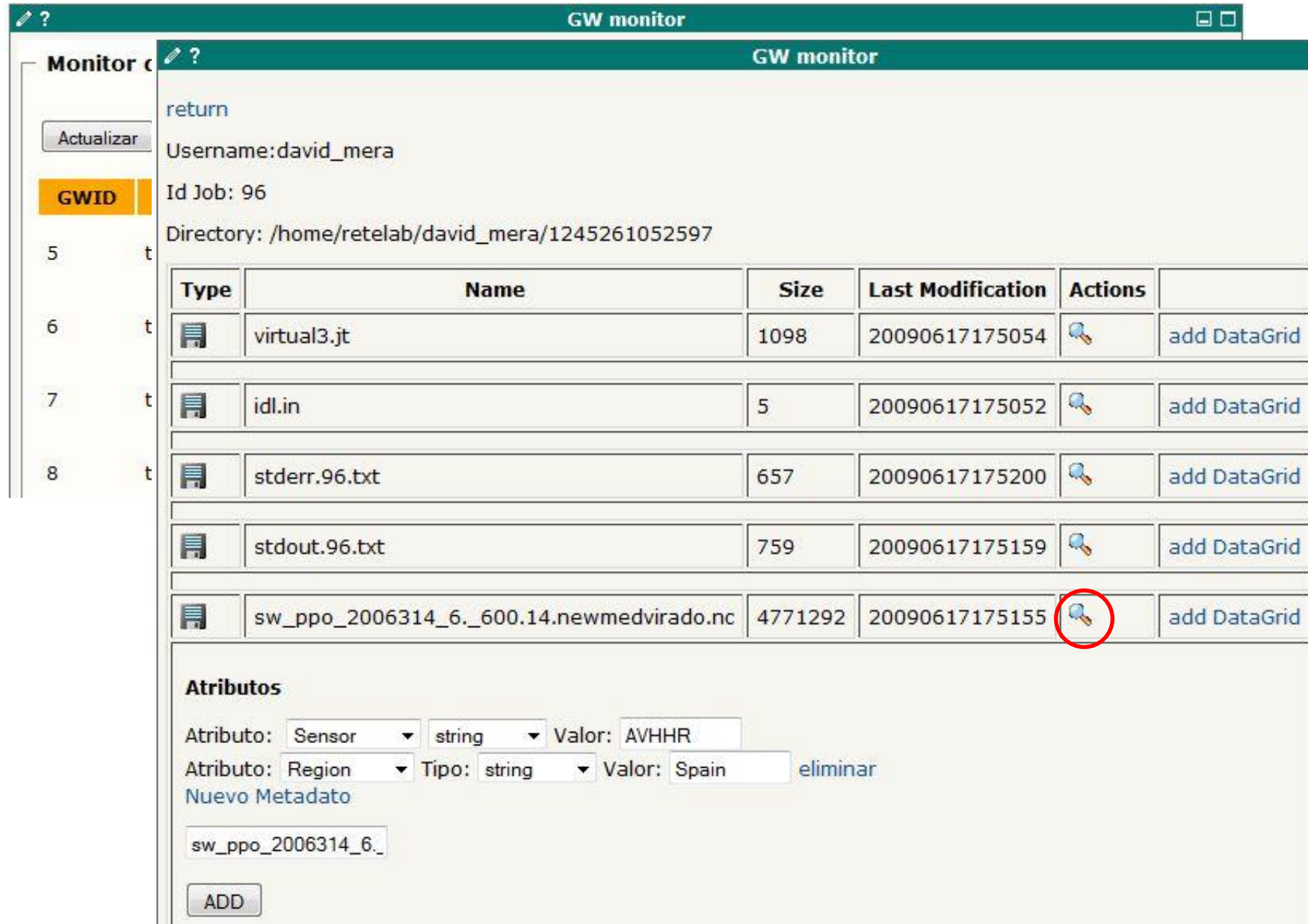
return
Username:david_mera
Id Job: 96
Directory: /home/retelab/david_mera/1245261052597

Type	Name	Size	Last Modification	Actions
virtual3.jt	virtual3.jt	1098	20090617175054	add DataGrid
idl.in	idl.in	5	20090617175052	add DataGrid
stderr.96.txt	stderr.96.txt	657	20090617175200	add DataGrid
stdout.96.txt	stdout.96.txt	759	20090617175159	add DataGrid
sw_ppo_2006314_6_600.14.newmedvirado.nc	sw_ppo_2006314_6_600.14.newmedvirado.nc	4771292	20090617175155	add DataGrid

Atributos

Atributo: Sensor string Valor: AVHHR
Atributo: Region Tipo: string Valor: Spain eliminar
Nuevo Metadato
sw_ppo_2006314_6_
ADD

Retelab Use case



The screenshot displays the 'GW monitor' application interface. On the left, a sidebar contains a 'Monitor' section with an 'Actualizar' button and a 'GWID' tab. The main content area shows a file list with columns for Type, Name, Size, Last Modification, and Actions. The file 'sw_ppo_2006314_6_.600.14.newmedvirado.nc' is highlighted, and its search icon is circled in red. Below the table, the 'Atributos' section allows for metadata management, including a form to add or edit attributes like 'Sensor' and 'Region'.

return
Username: david_mera
Id Job: 96
Directory: /home/retelab/david_mera/1245261052597

Type	Name	Size	Last Modification	Actions
	virtual3.jt	1098	20090617175054	add DataGrid
	idl.in	5	20090617175052	add DataGrid
	stderr.96.txt	657	20090617175200	add DataGrid
	stdout.96.txt	759	20090617175159	add DataGrid
	sw_ppo_2006314_6_.600.14.newmedvirado.nc	4771292	20090617175155	add DataGrid

Atributos

Atributo: Valor:

Atributo: Tipo: Valor: [eliminar](#)

[Nuevo Metadato](#)

Retelab Use case

The screenshot displays a web application interface with a sidebar on the left and a main content area. The sidebar contains a 'Monitor' section with a table of data. The main content area shows a 'GW monitor' window with a dialog box open. The dialog box is titled 'Abriendo loaderdavid_meraexample2.nc.jnlp' and contains the following text: 'Ha escogido abrir loaderdavid_meraexample2.nc.jnlp el cual es un: Java Web Start de: http://...'. Below this text, there is a question: '¿Qué debería hacer Firefox con este archivo?'. There are three radio button options: 'Abrir con' (selected), 'Guardar archivo', and 'Hacer esto automáticamente para los archivos como éste de ahora en adelante.'. The 'Abrir con' option has a dropdown menu showing 'javaws'. At the bottom of the dialog are 'Cancelar' and 'Aceptar' buttons.

Monitor

Actualizar

GWID

5	t	
6	t	
7	t	
8	t	

GW monitor

return

Username

Id Job: 9

Directory

Type

sw_ppo_2006314_6_600.14.newmedvirado.nc | 4771292 | 20090617175155 | add DataGrid

Atributos

Atributo: Sensor string Valor: AVHHR

Atributo: Region Tipo: string Valor: Spain eliminar

Nuevo Metadato

sw_ppo_2006314_6_

ADD

Retelab Use case

The screenshot displays the Retelab GW monitor interface, which is divided into several panels:

- Monitor Panel (Left):** Contains a table with columns for 'Type' and 'Atributos'. The 'GWID' column is highlighted in orange. Below the table is an 'Actualizar' button and a section for 'Atributos' with dropdown menus for 'Sensor' and 'Region', and a text input field containing 'sw_ppo_2006314_6_'. An 'ADD' button is located at the bottom.
- Metadata Editor Panel (Middle):** A dialog box titled 'Unidata IDV - Map View - One Pane' is open. It shows a file named 'loaderdavi' and a URL 'http://...'. The '¿Qué debería hacer?' section has three options: 'Abrir con' (selected), 'Guardar a', and 'Hacer est...'. Below this is a text input field containing 'sw_ppo_2006314_...'. The 'Atributos' section is also visible, with 'Sensor' and 'Region' dropdowns and a 'Tip' dropdown.
- Map View Panel (Right):** A map titled 'Unidata IDV - Map View - One Pane' is displayed. The map shows a contour plot with a red line and a blue area. The map is titled 'prn_prod - Contour Plot View'. The status bar at the bottom of the map view shows 'Memory: 96/136/533 MB' and 'Latitude: 37,9 Longitude: -22,2 A...'. The map view has a menu bar with 'File', 'Edit', 'Displays', 'Data', 'Tools', and 'Help', and a toolbar with various icons.

Index

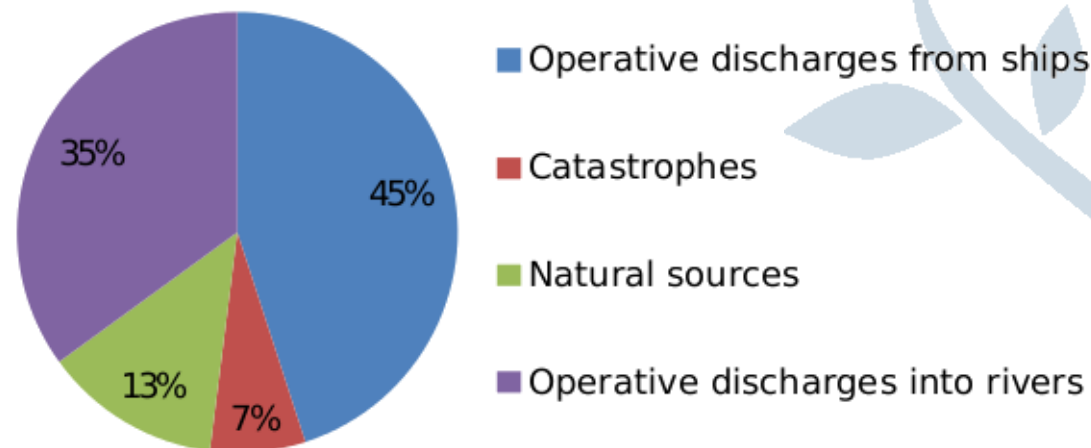
1. Context and motivation
2. Main goals
3. Retelab project - Virtual laboratory development
4. **Sentinazos project - Virtual laboratory validation**
5. On going work



Sentinazos

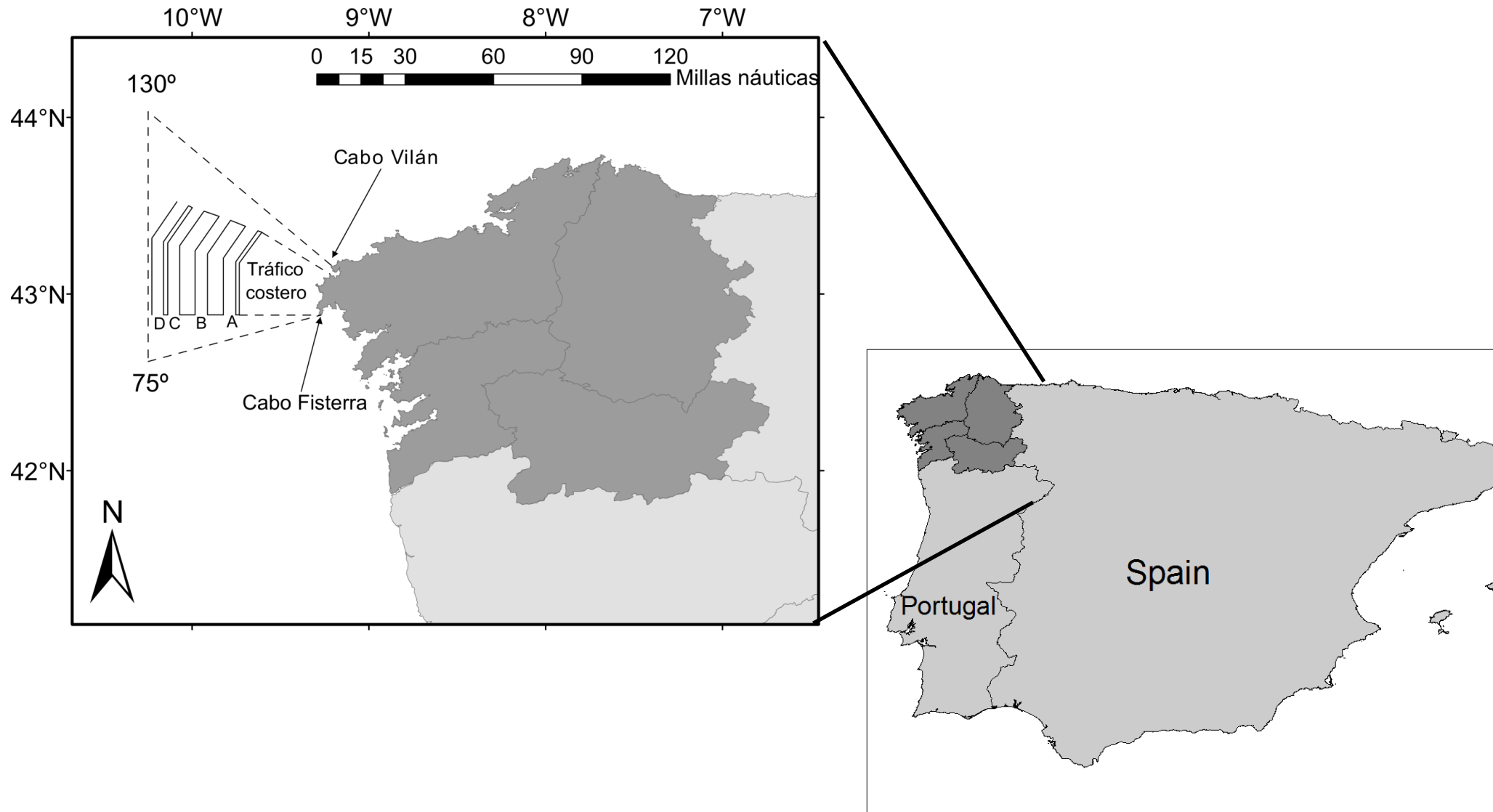
Introduction

- The international trade is mainly supported by maritime transport
- The intensive traffic sails along the Exclusive Economic Zones (EEZ) of countries and generates important pollution problems.
- Only the 7% of oil spills come from catastrophes like tanker and oil platform accidents.

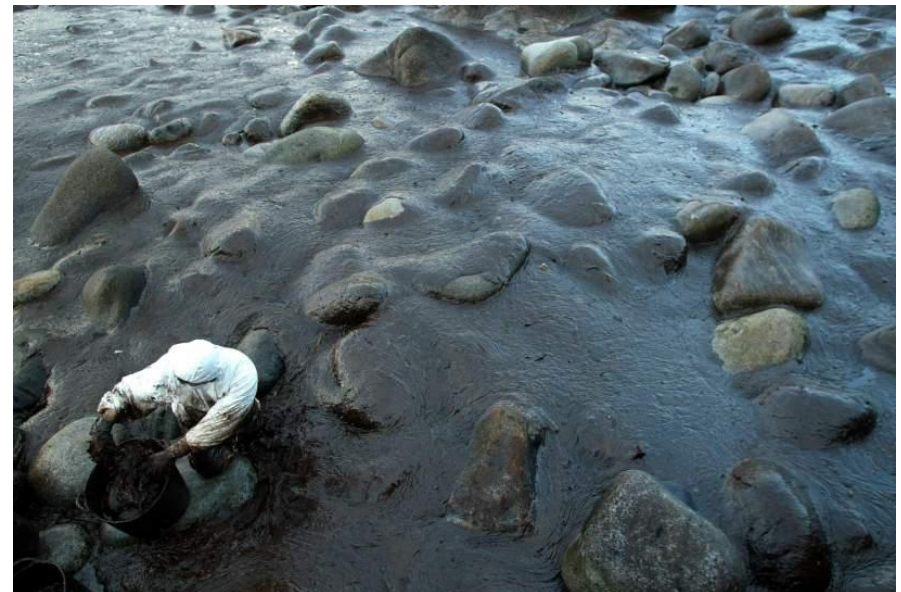


Sentinazos

Introduction



Sentinazos Introduction



Sentinazos

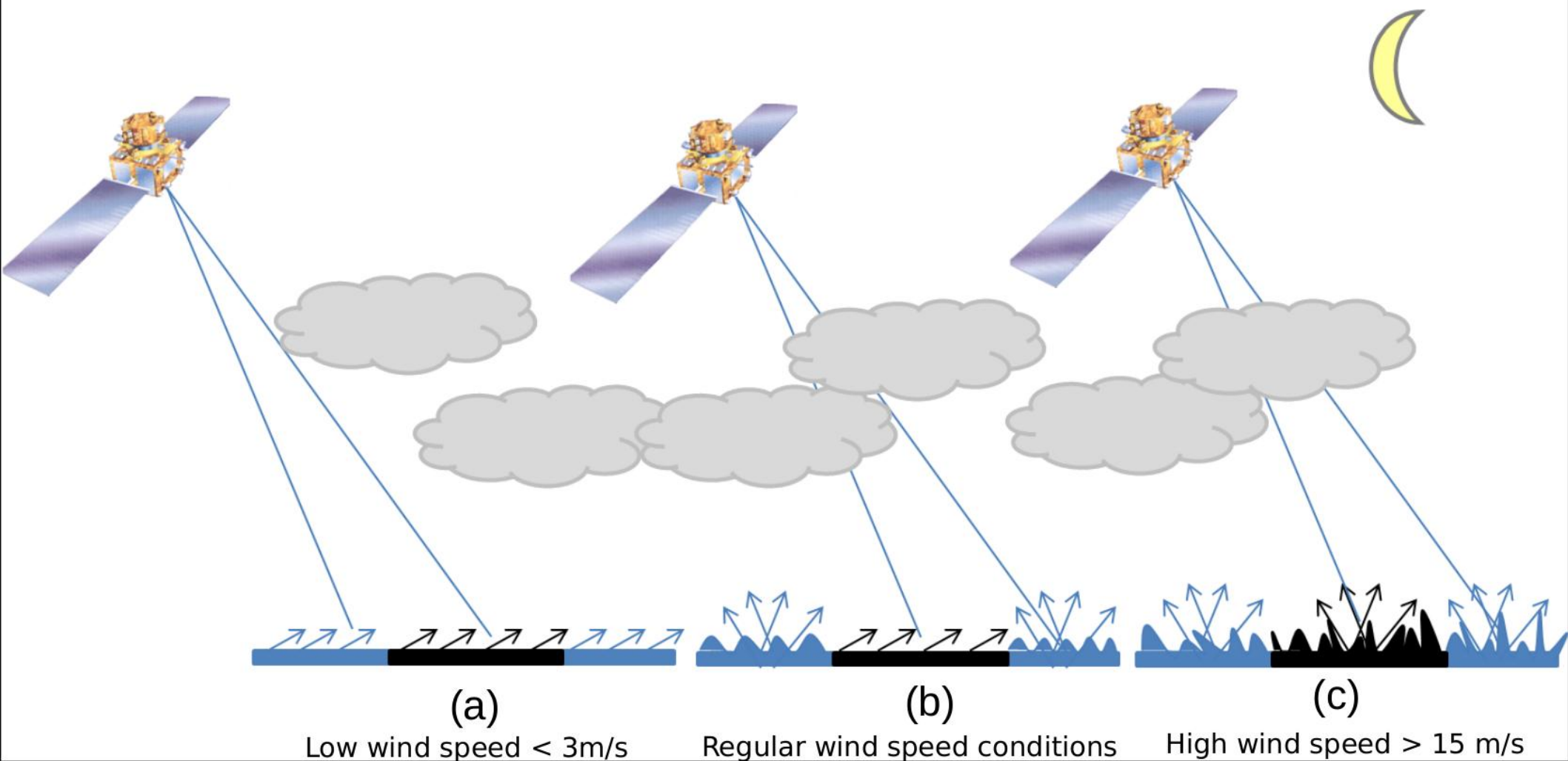
Introduction



Sentinazos

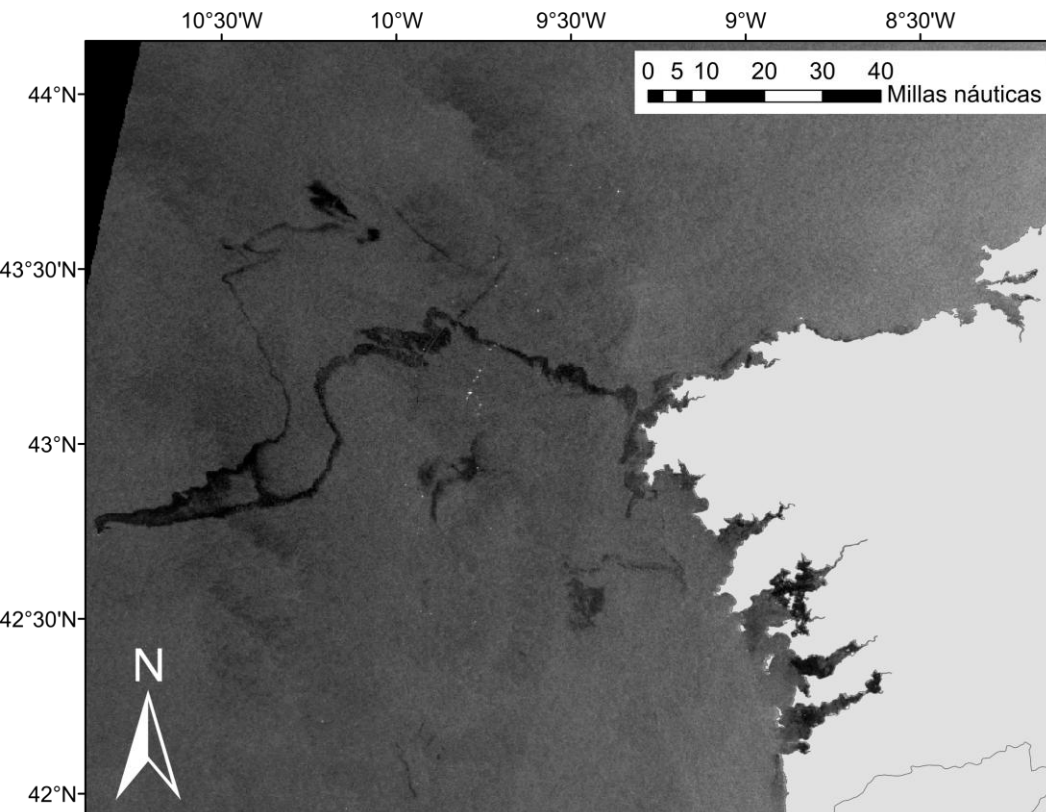
Introducción

- Synthetic Aperture Radar

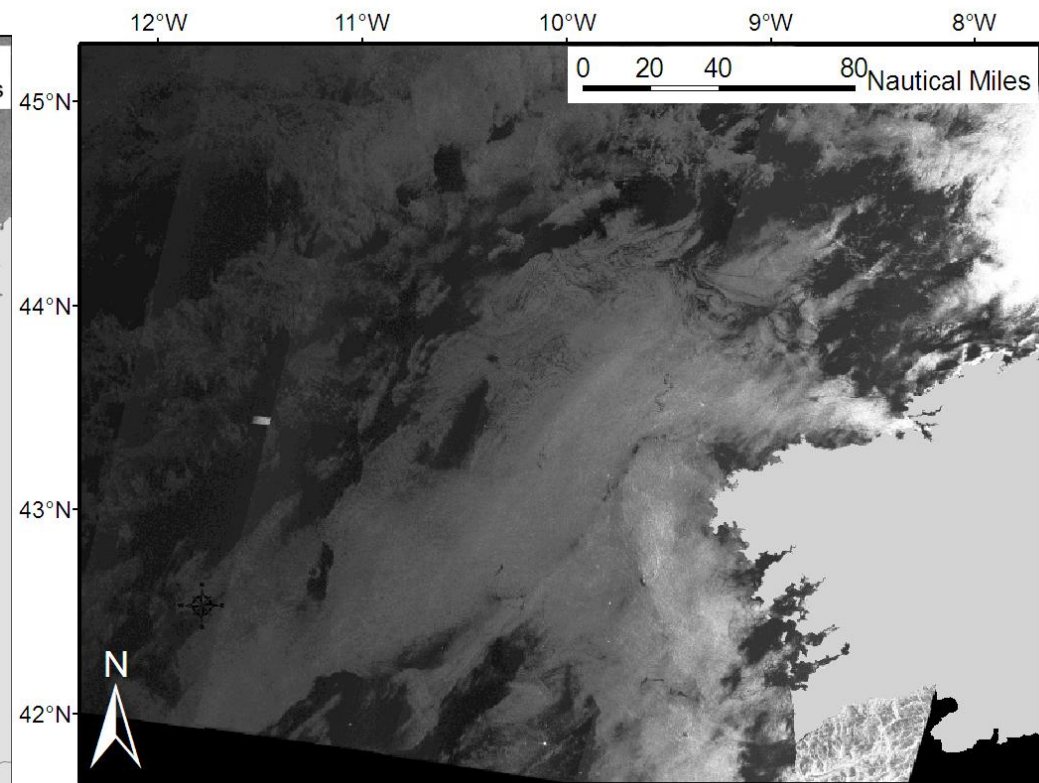


Sentinazos Introduction

- Synthetic Aperture Radar - Examples



Prestige catastrophe
(Envisat, 17/11/2002)



Galician coast
(Envisat, 13/10/2008)

Sentinazos

Goal

Hypothesis

1. Wind data could be used to improve the SAR image segmentation
2. Oil spill shape could be relevant to correctly classify them

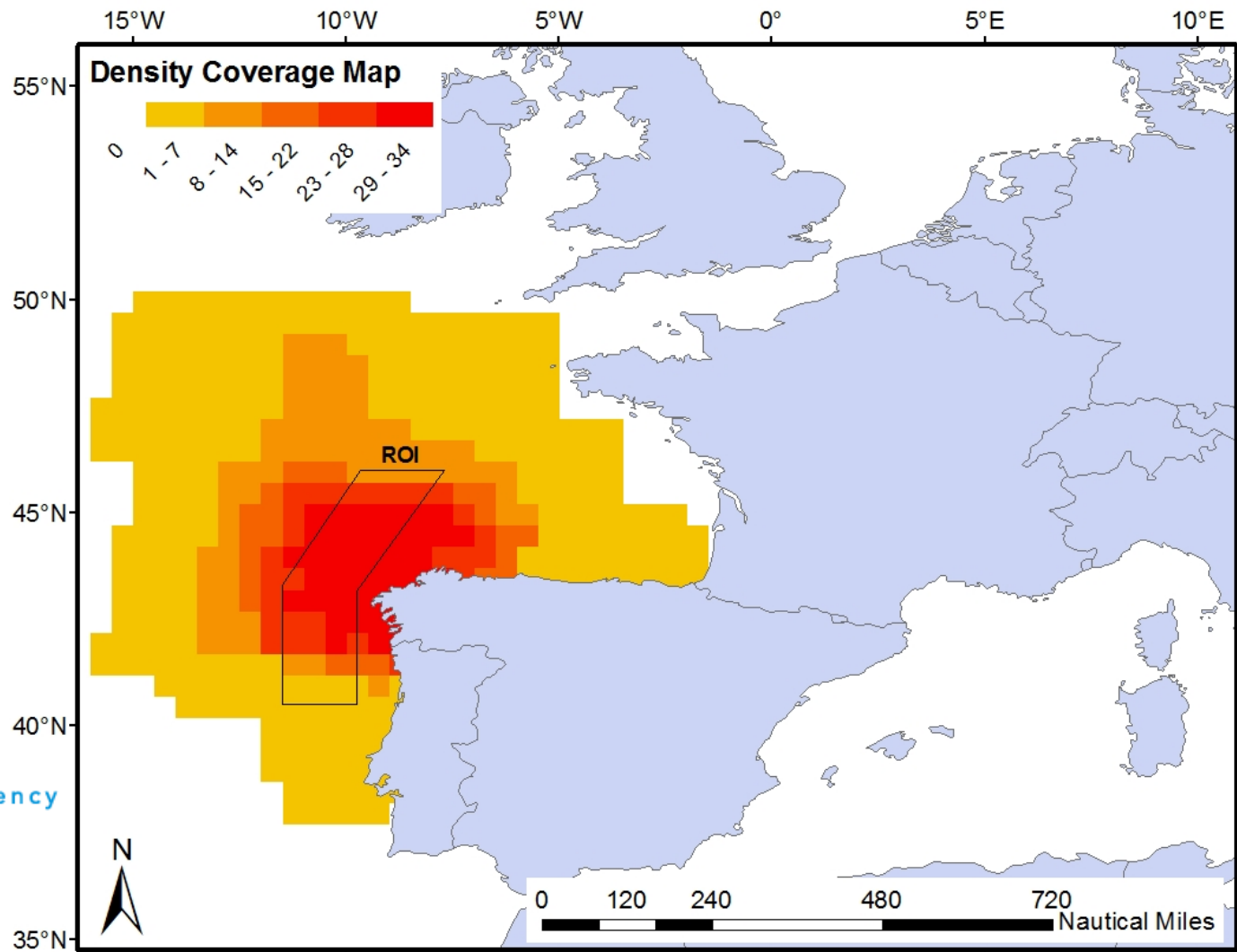
Goal

- To develop an automatic oil spill detection system based on SAR images and focused on the galician coast. This system should take advantages of win data as well as candidate shape features to improve detection results

Sentinazos

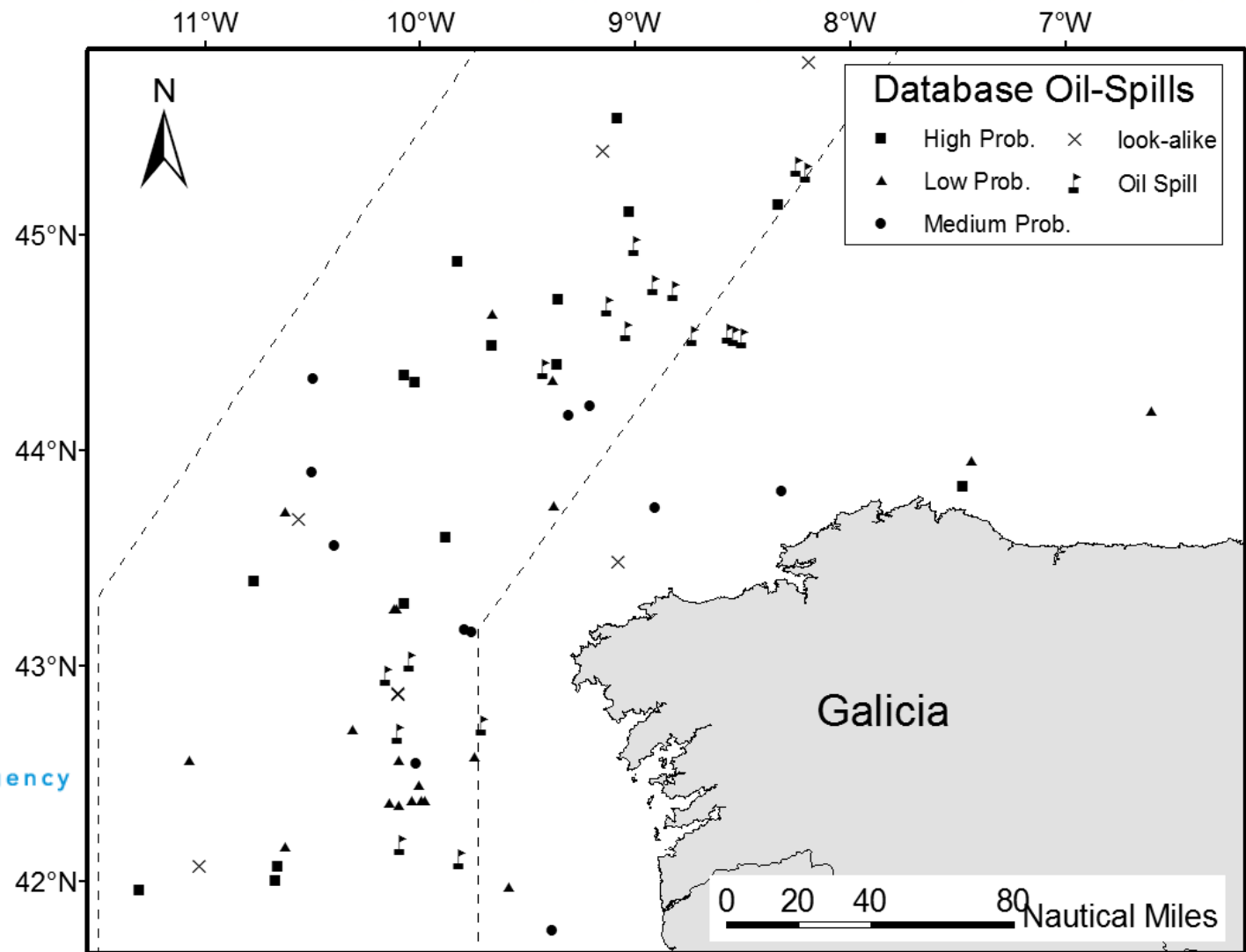
Methodology

- Dataset- 47 SAR images from the Envisat



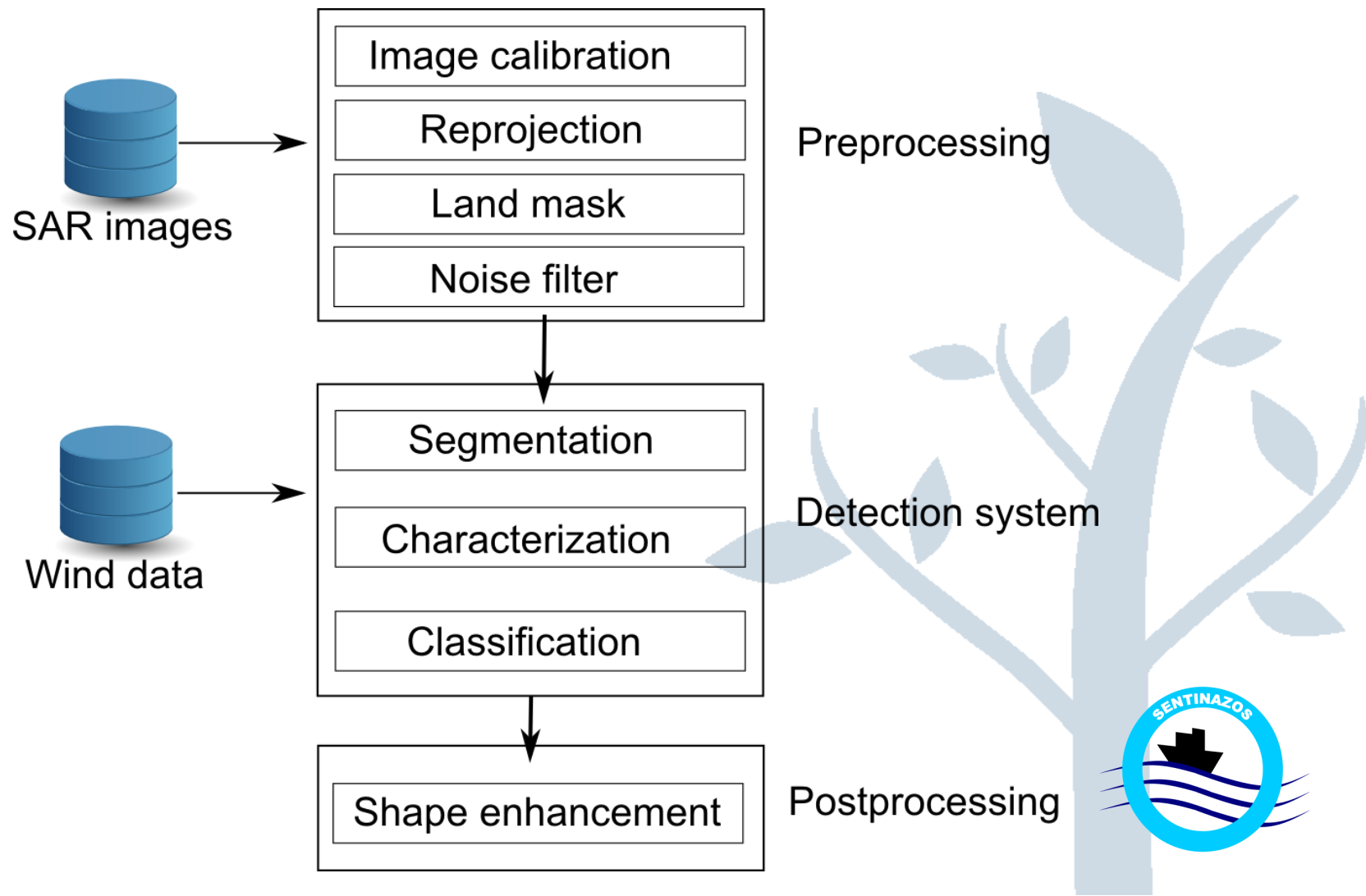
Sentinazos Methodology

- Dataset- 47 SAR images from the Envisat



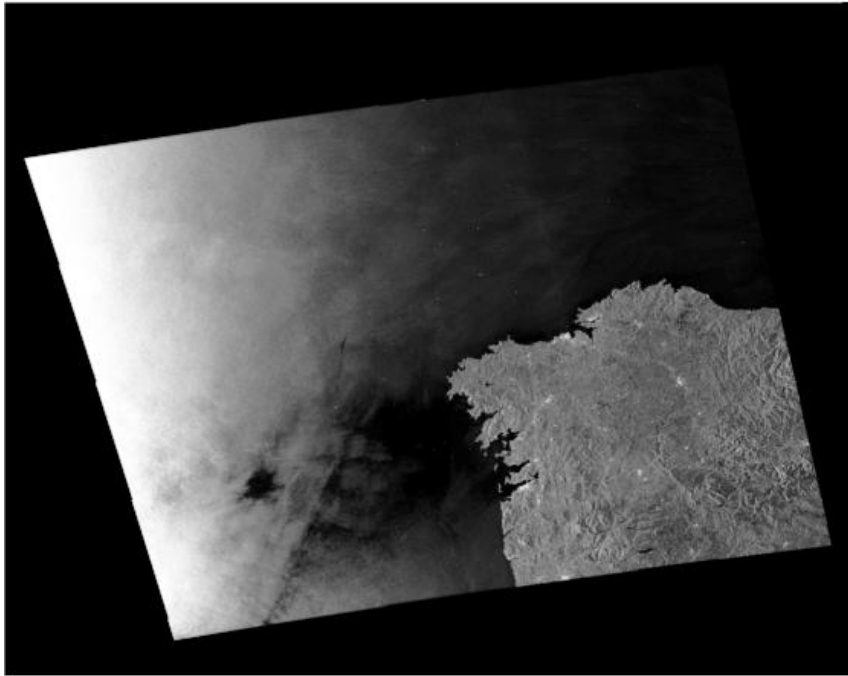
Sentinazos Methodology

- System architecture

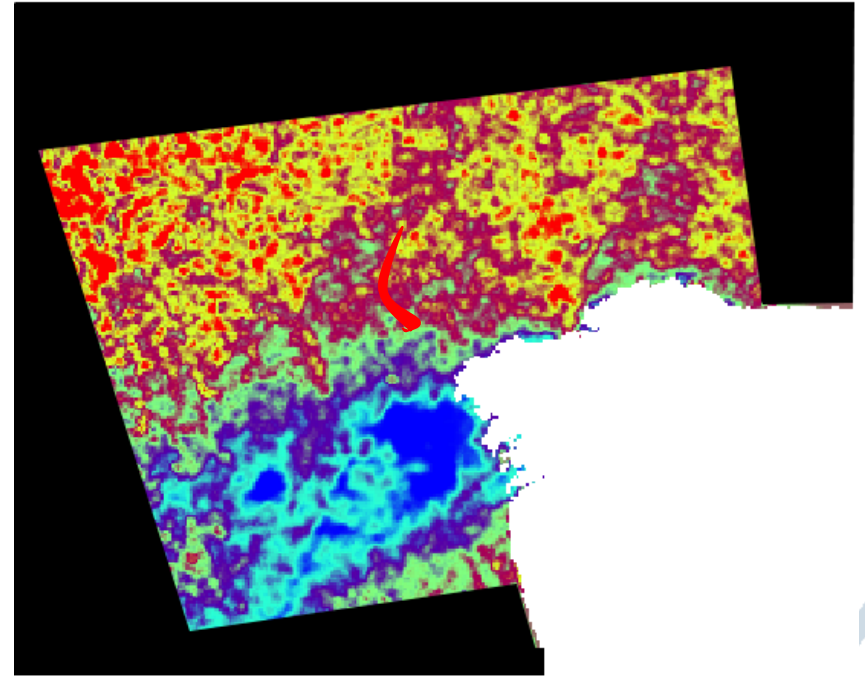


Sentinazos

Methodology - Segmentation



SAR image. Envisat, 01/06/2007

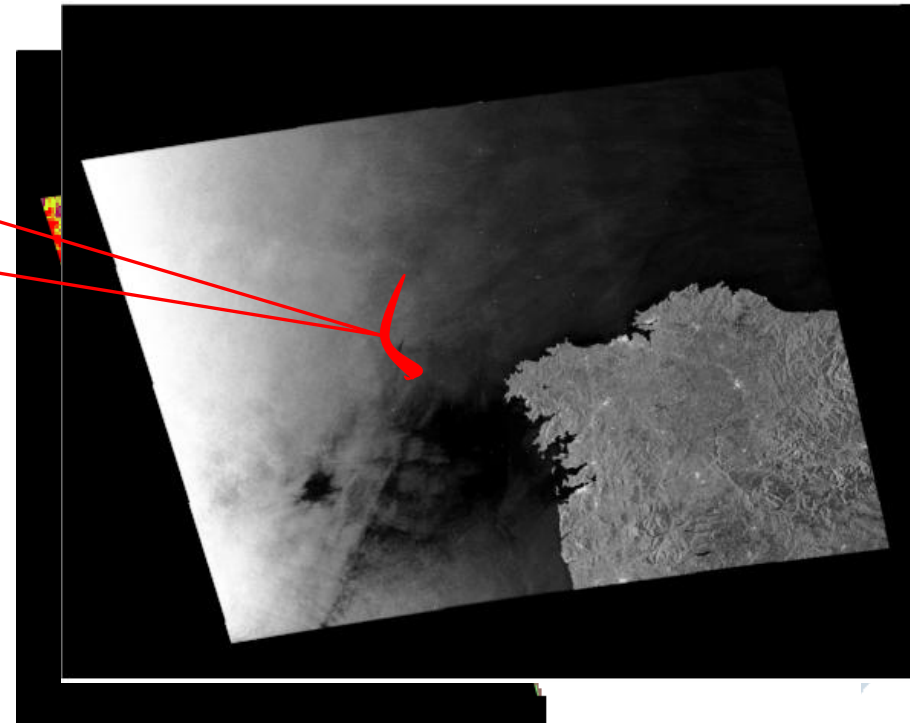


Wind intensity data. CMOD5 model

Sentinazos

Methodology - Segmentation

X	Y	Wind	I.A.	Intensity
2003	1212	5,4	41,3	0,021
1233	5298	3,2	23,74	0,5421
6832	4523	3,8	27,21	0,00234

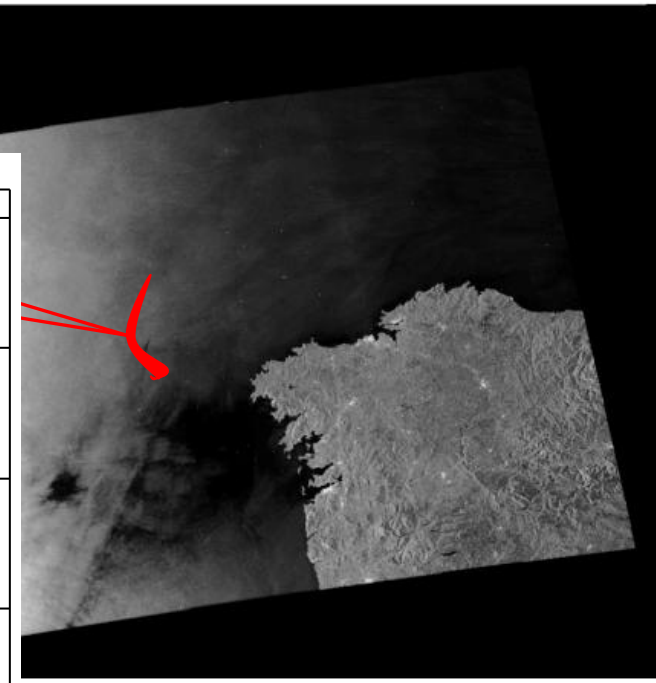
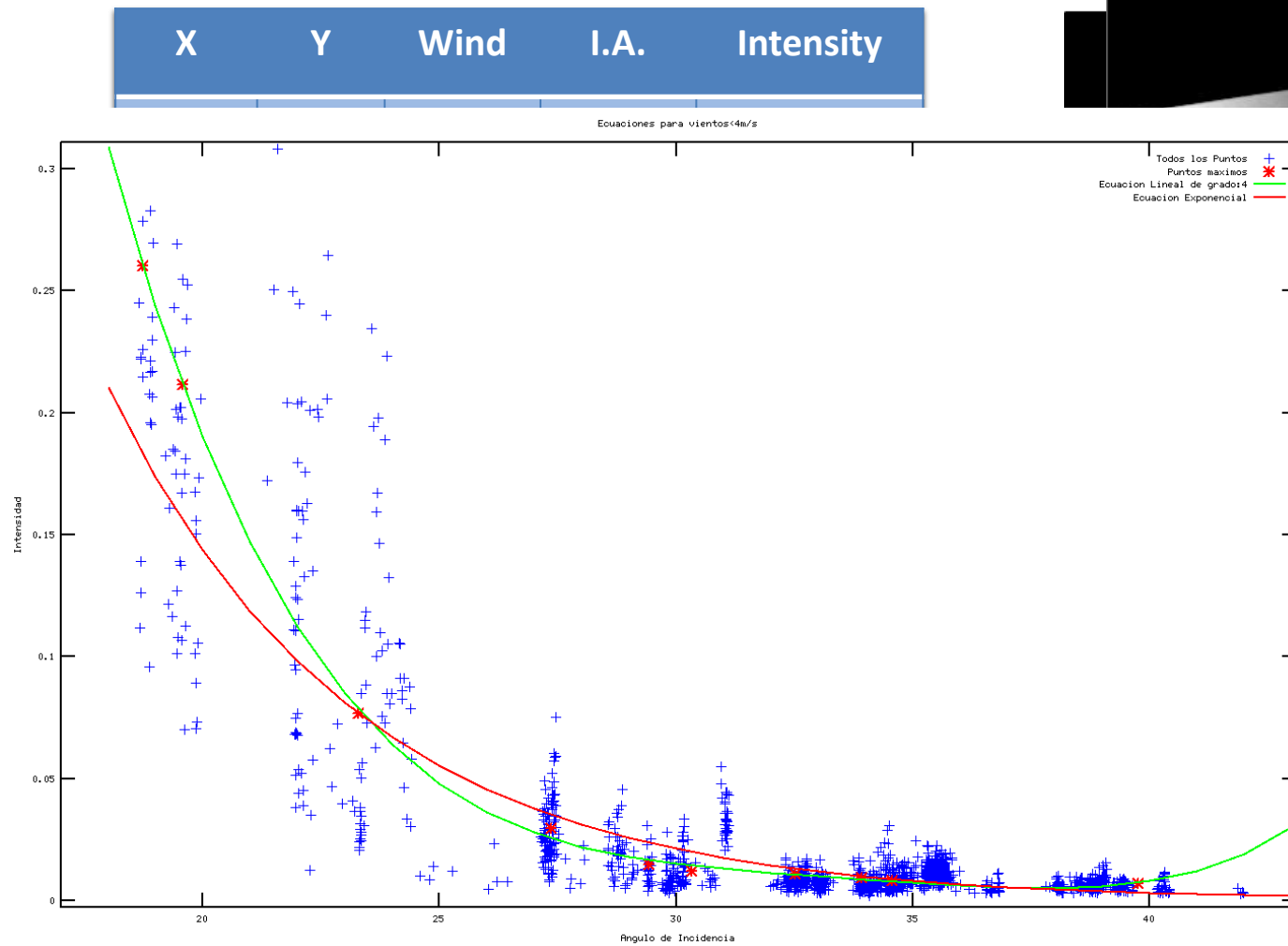


SAR image. Envisat, 01/06/2007

Wind intensity data. CMOD5 model

Sentinazos

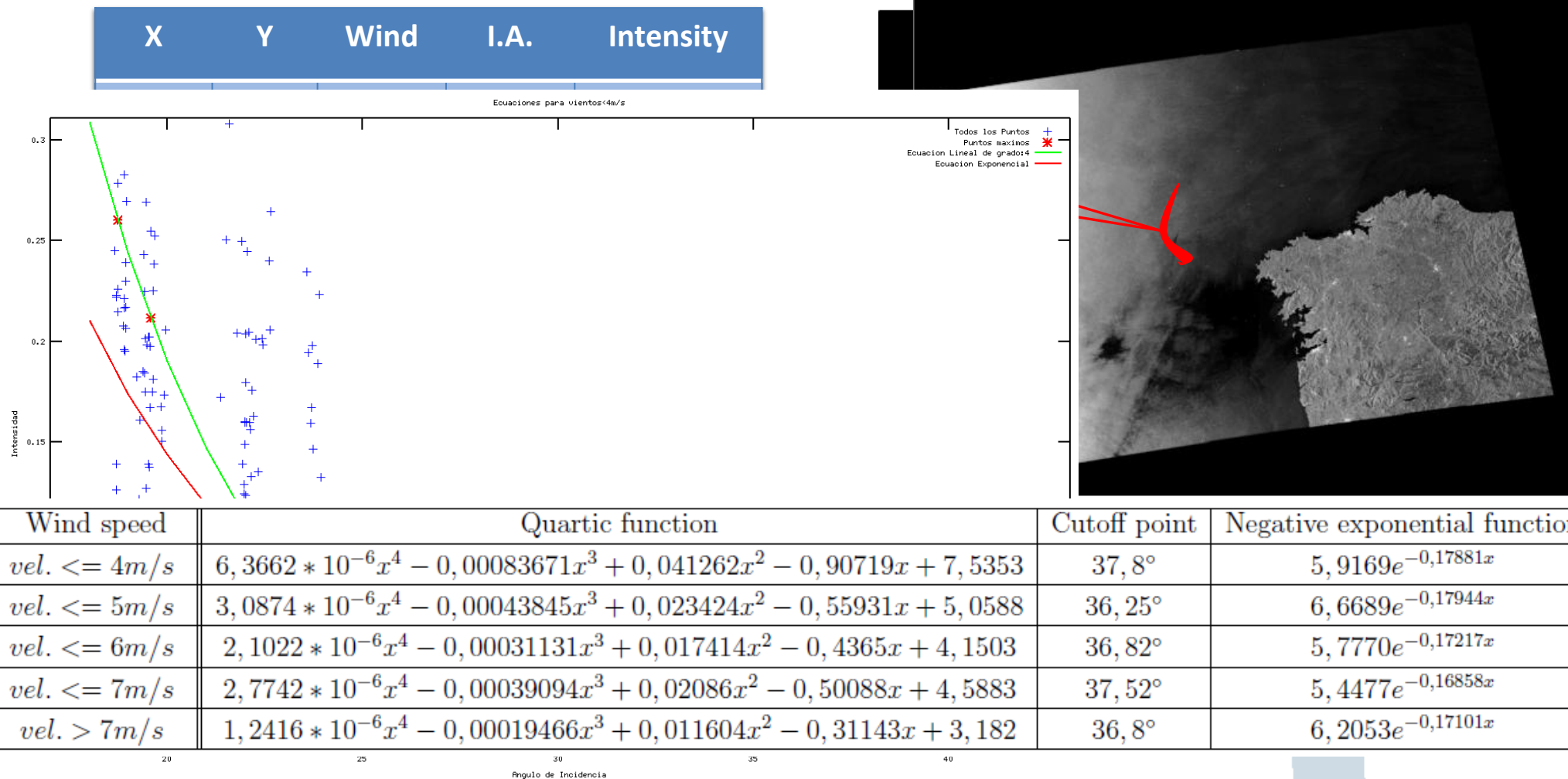
Methodology - Segmentation



ensity data. CMOD5 model

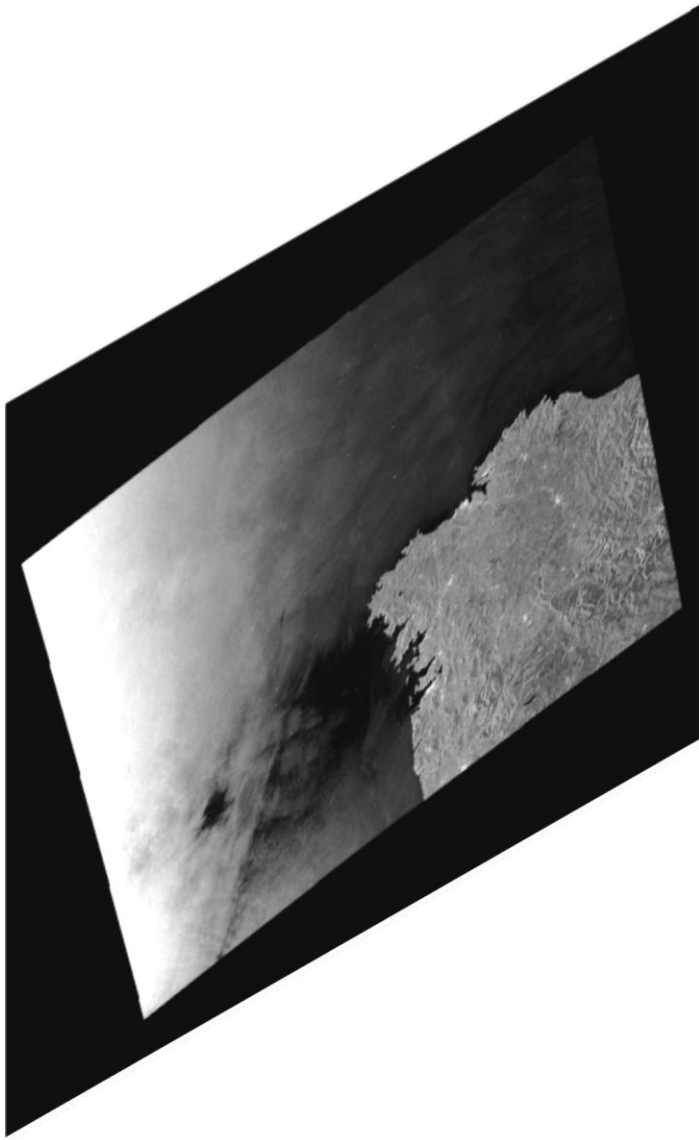
Sentinazos

Methodology - Segmentation



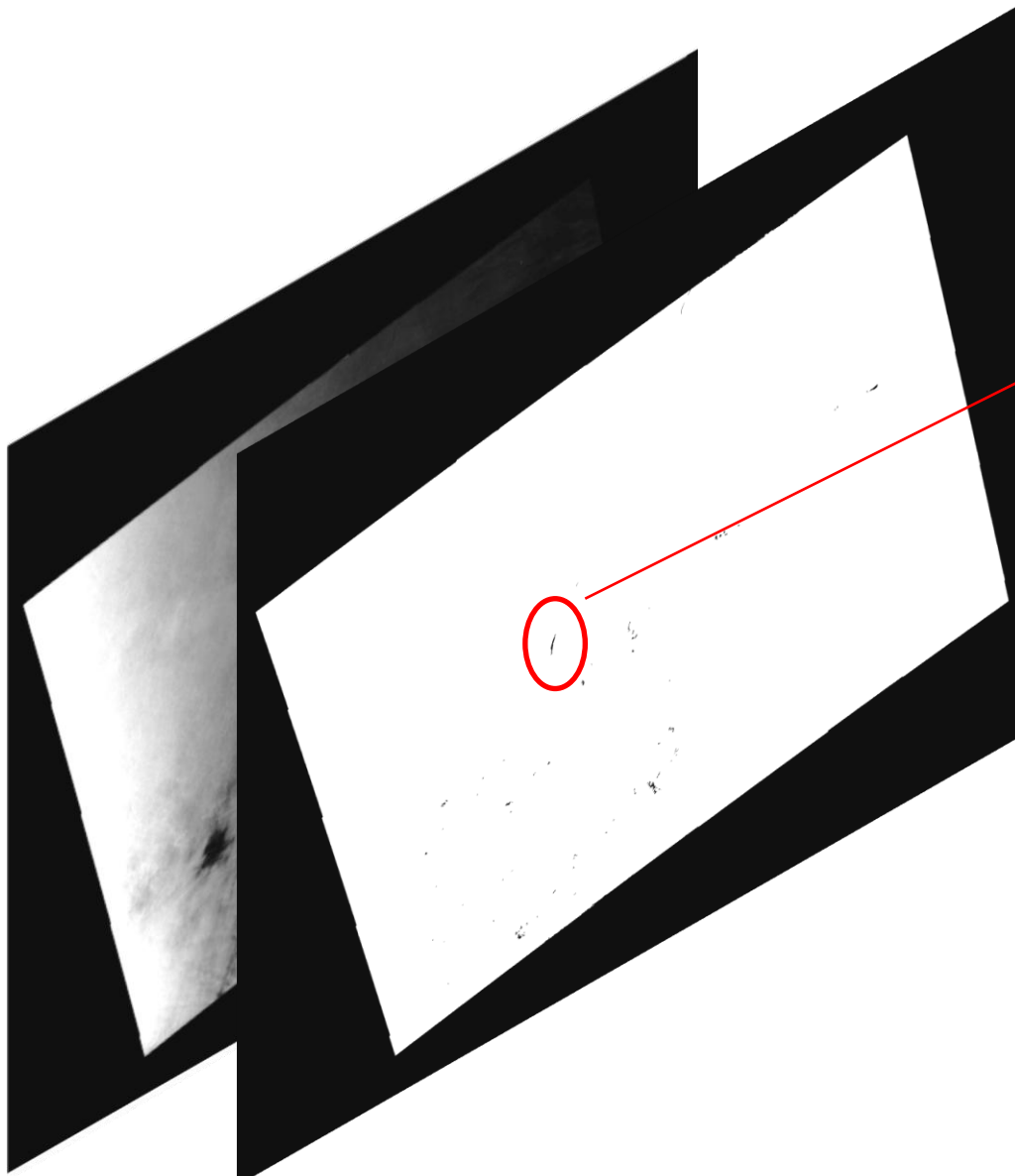
Sentinazos

Methodology - Characterization

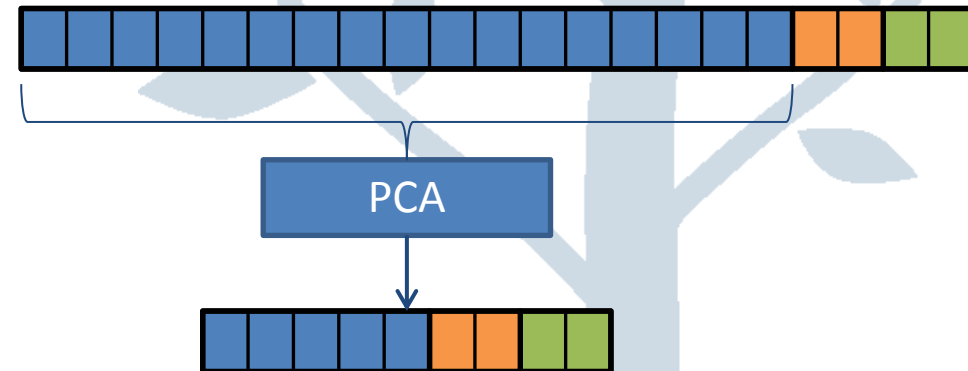


Sentinazos

Methodology - Characterization



- 17 shape characteristics
- 2 physical characteristics
- 2 contextual characteristics



Sentinazos

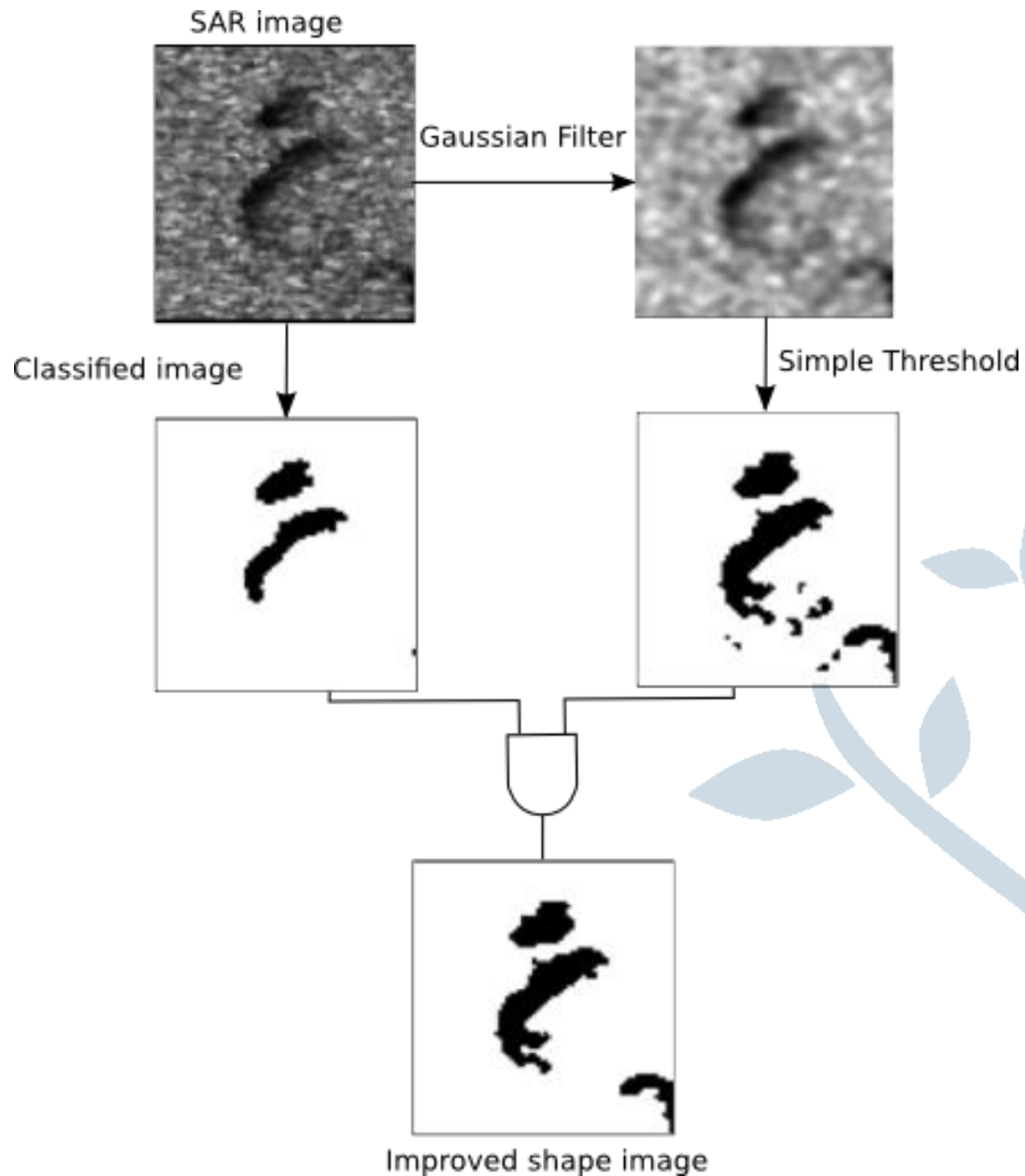
Methodology - Classification

- Classification
 - Clustering of oil spills and look alike.
 - Evaluation of the characteristics vector.
 - Developed classifiers
 - Artificial Neural Network
 - Decision Tree



Sentinazos

Methodology - Postprocessing



Sentinazos Results

	Validation subset		Test subset	
	Sentinazos	False positives	Sentinazos	False positives
ANN	85,7%	85,2%	92,9%	96,3%
Decision tree	92,9%	85,2%	92,9%	92,6%



Parallel version	HW characteristics	Segmented pixels	Segmented candidates	Processing time	Improvement
OpenMP + TBB	Intel Core 2 Duo Processor E6400 (2.13GHz), 3GB RAM	210064	670	45,33 sg.	27,07%
		36366	108	18,57 sg.	23,89%
		27211	65	15,29 sg.	26,81%

Sentinazos Results

- Desktop application

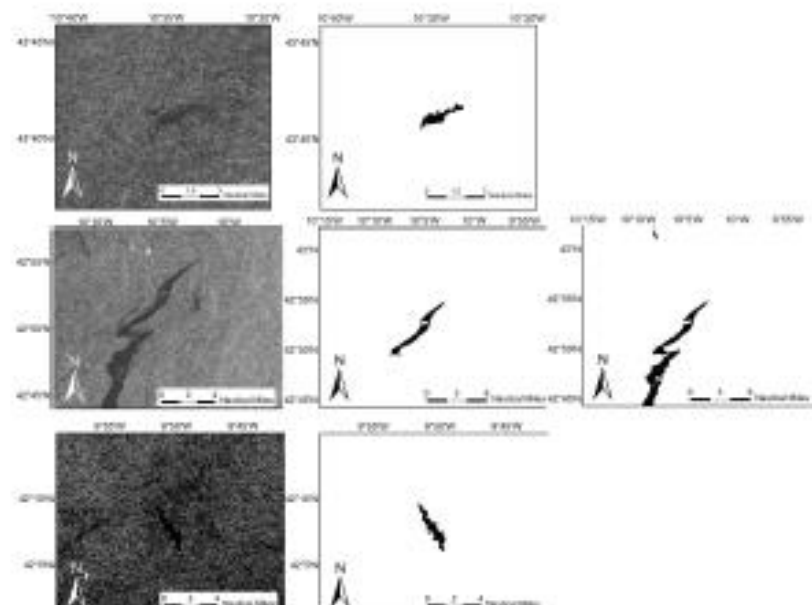
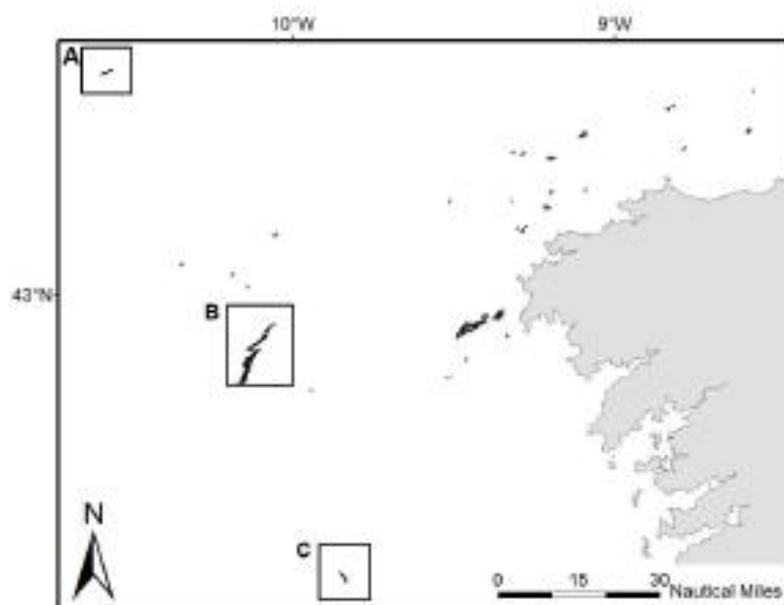
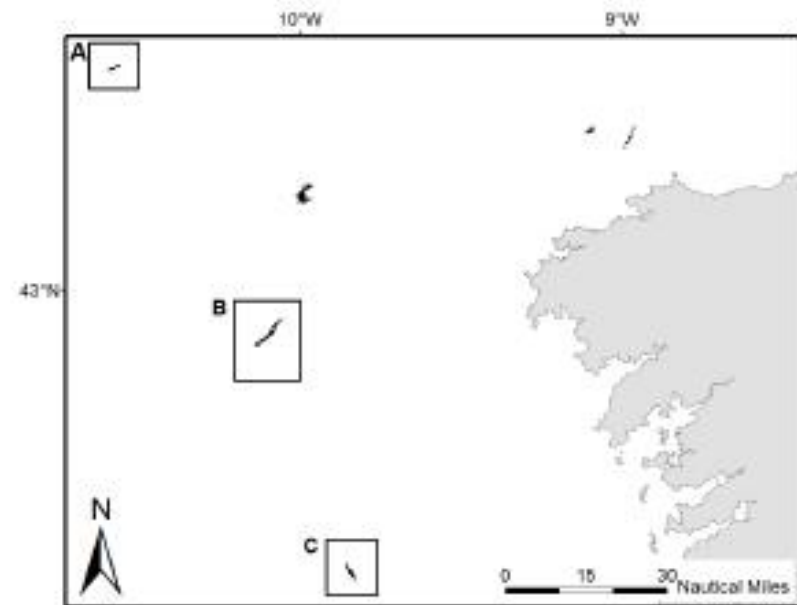
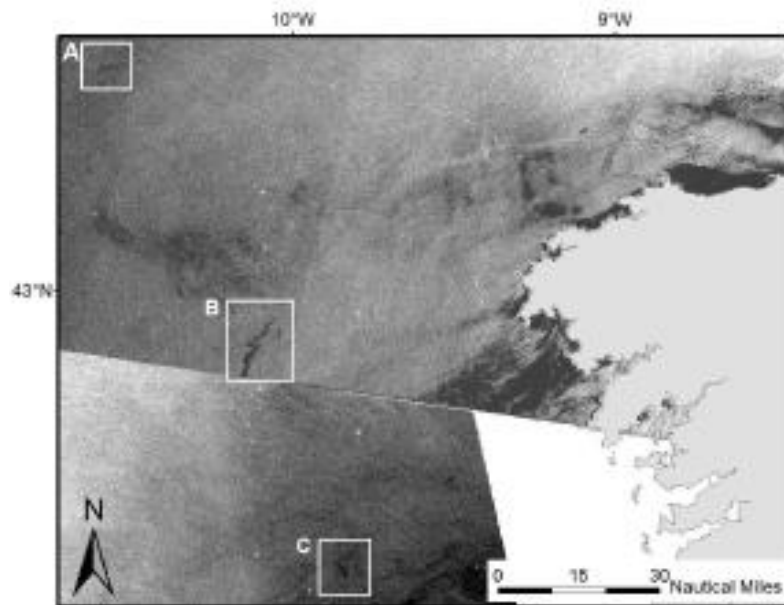


Sentinazos Results

- Portlet Grid

The screenshot displays the RETELAB Grid Portal interface. At the top, the browser shows the URL "RETELAB Grid Portal". The page header includes the RETELAB logo and the text "LABORATORIO VIRTUAL PARA LA RED NACIONAL DE TELEDETECCIÓN OCEANOGRÁFICA". A navigation menu contains links for "Welcome", "Administration", "file-manager", "Primary Production", "jobsubmit-portlet", "Proxymanager Portlet", "sentinazos", and "gridway-datagrid-portlet". The "sentinazos" portlet is active, showing a form for configuring a project. The form includes fields for "Título del trabajo:" (galicia12-11-2012), "Acción a realizar:" (Segmentar y Clasificar), and "Clasificador utilizado:" (Redes Neuronales). There are also fields for "Imagen SAR:", "IA:", "Vientos:", "Remoto:", and "Local:". A "Enviar" button is located at the bottom right of the form. On the right side, there is a "Virtual Data Base" section with dropdown menus for "Region", "Sensor", and "Parameter", a "Search" button, and a file selection area with a "Select input files" button. Below this, a list of files is shown: "SARimage.tif", "SARimageWind.nc", and "SARimageIA.tif". The bottom of the page features a logo for "SENTINAZOS UDC · UDC".

Sentinazos Results



Future

- Sentinazos
 - New SAR data sources
 - Sentinel-1. Currently, the Envisat is not operative
 - Integration of new remote sensing sources
 - Buoys
 - Optical sensors,
 - Aircraft missions - SLAR sensor
 - ...
 - Algorithm improvement
 - Contextual data
 - New wind model
 - Multi classifier systems



Thank you

Development, deployment and validation of an oceanographic
virtual laboratory based on Grid computing