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

David Mera Pérez

Brno, Czech Republic, October 21th 2013

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Instituto de Investigacións Tecnolóxicas

CiTiUS Centro Singular de Investigación en Tecnoloxías da Información

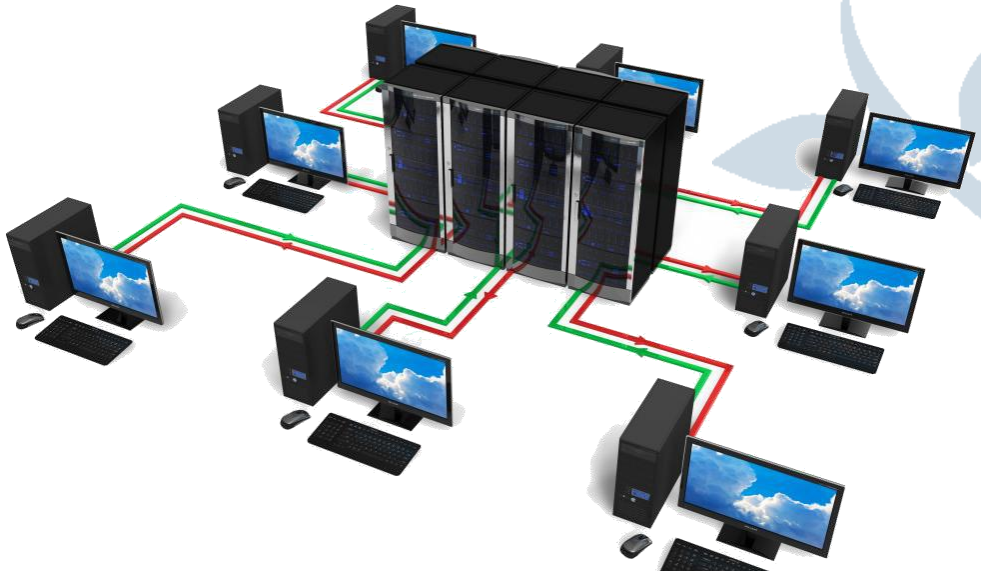
Index

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

Index

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 Standars
  - Virtual Organizations management

Index

Main goals

- To develop a user-friendly distributed computational environment based on Grid computing.

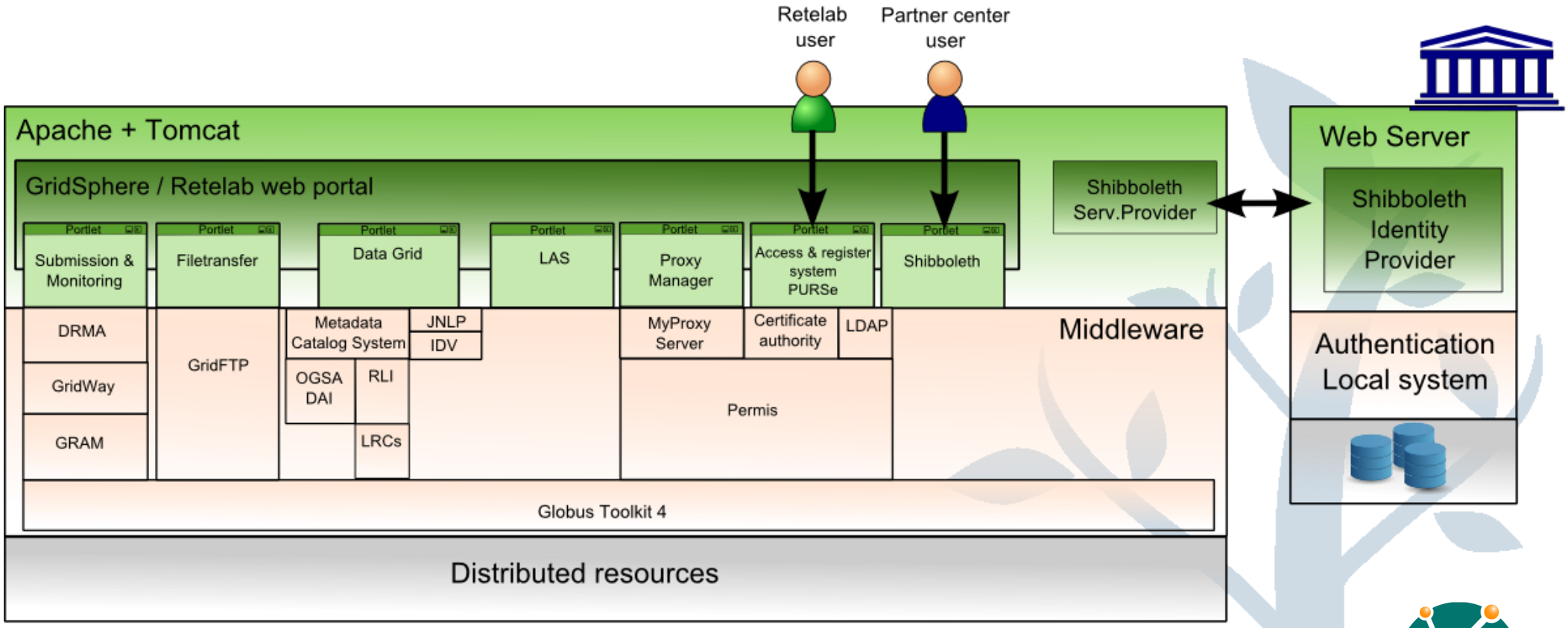
- To develop an oceanographic application to test the Grid environment.
  - An oil spill automatic detection system based on the analysis of satellite Synthetic Aperture Radar imaging.

Index

Retelab Basis

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

Retelab Architecture



RETELAB

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

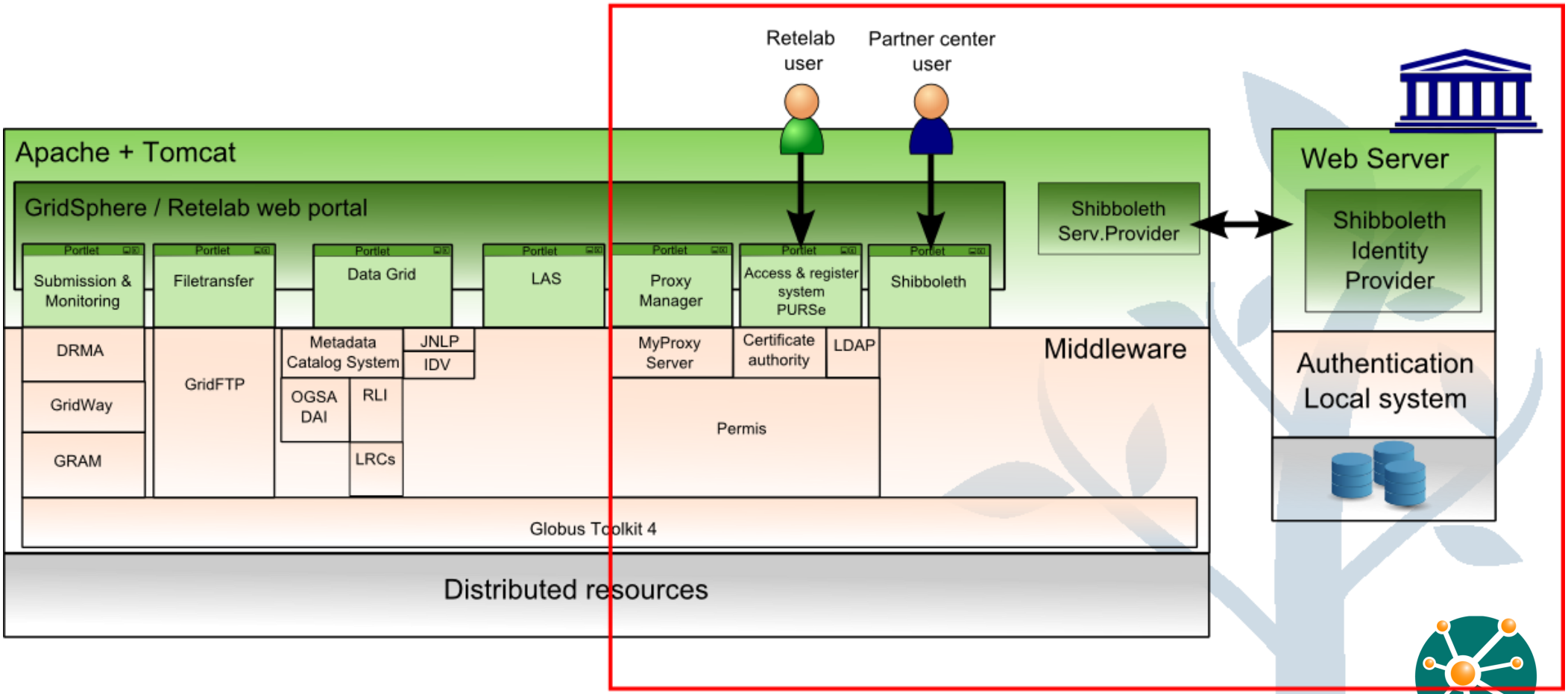
vs

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

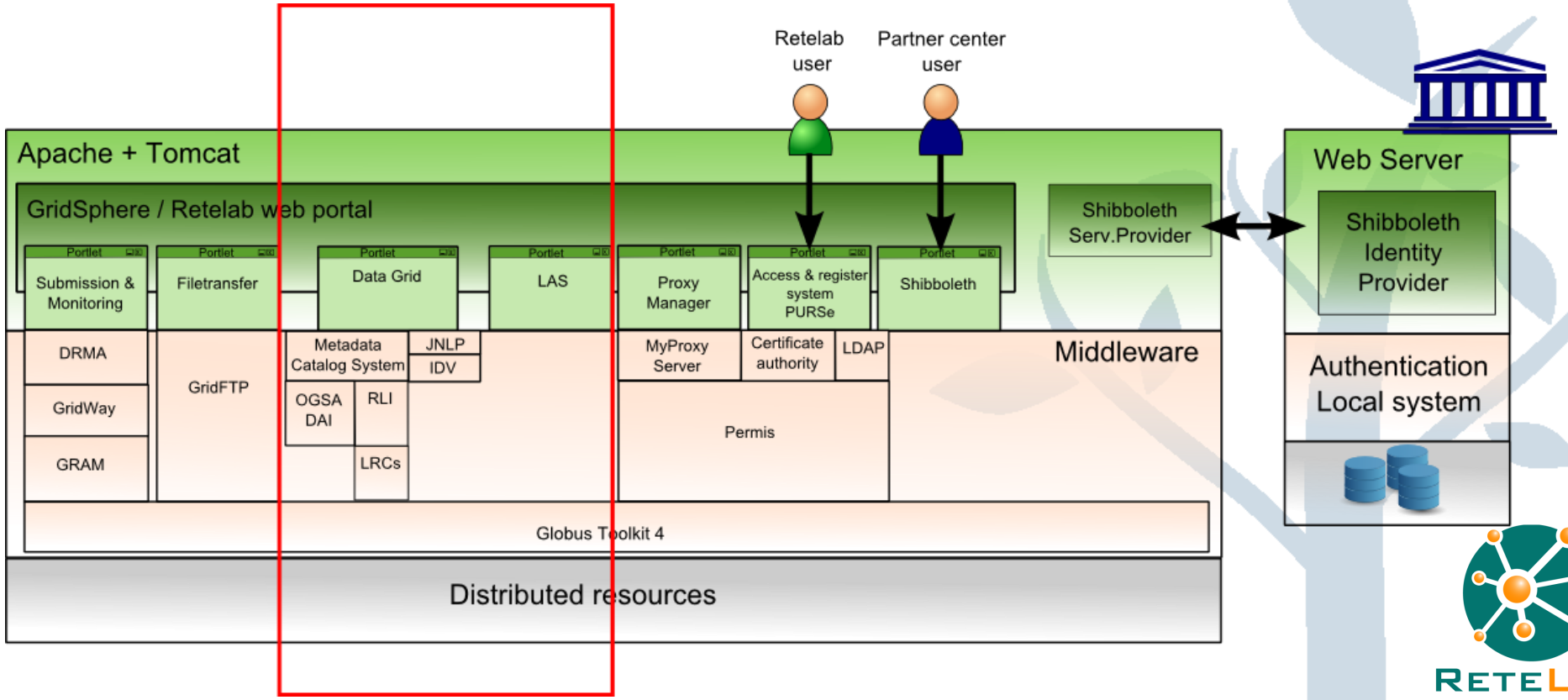
Retelab User access and registration system

- Arquitectura



Retelab Distributed storage system

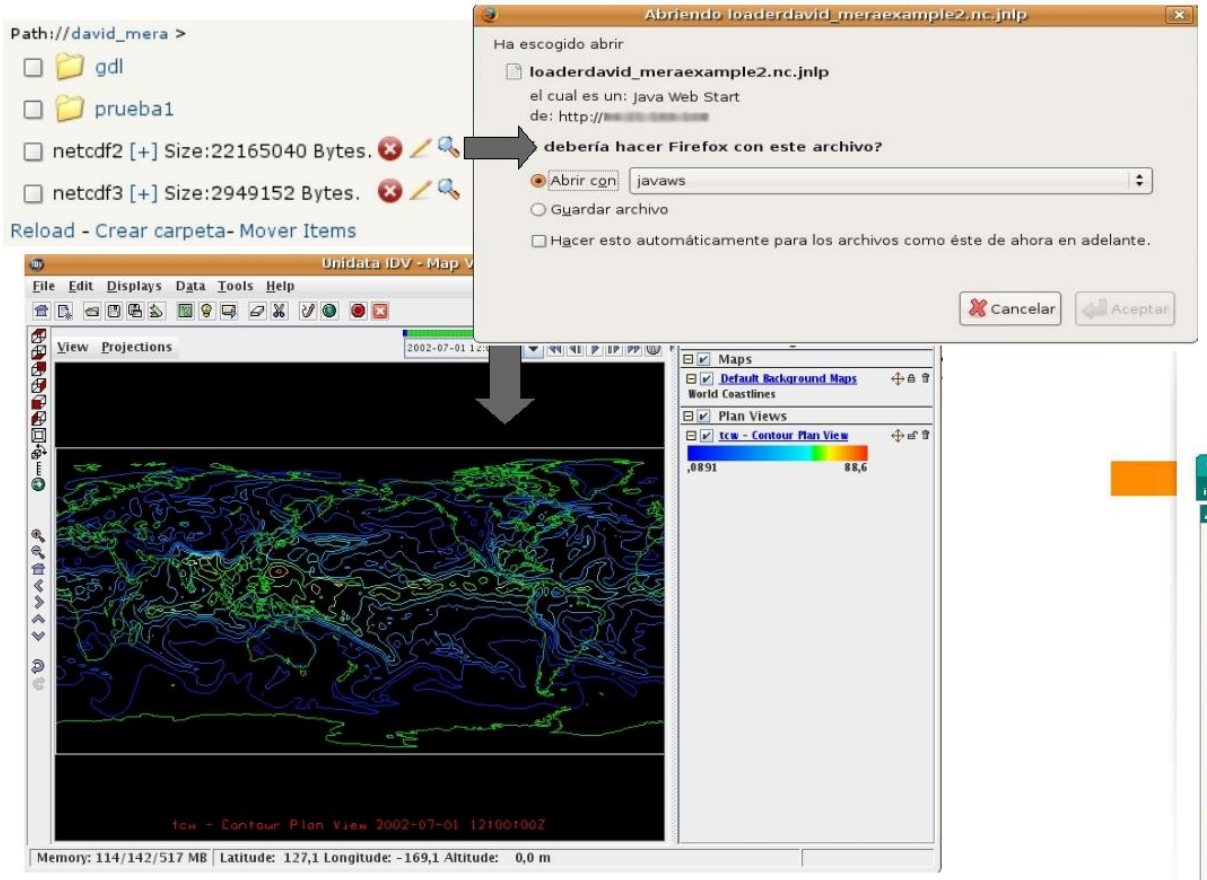
- Web interface integration
- Based on Metadata - ISO 19115, Geographic Information
- Integration of visualization tools



RETELAB

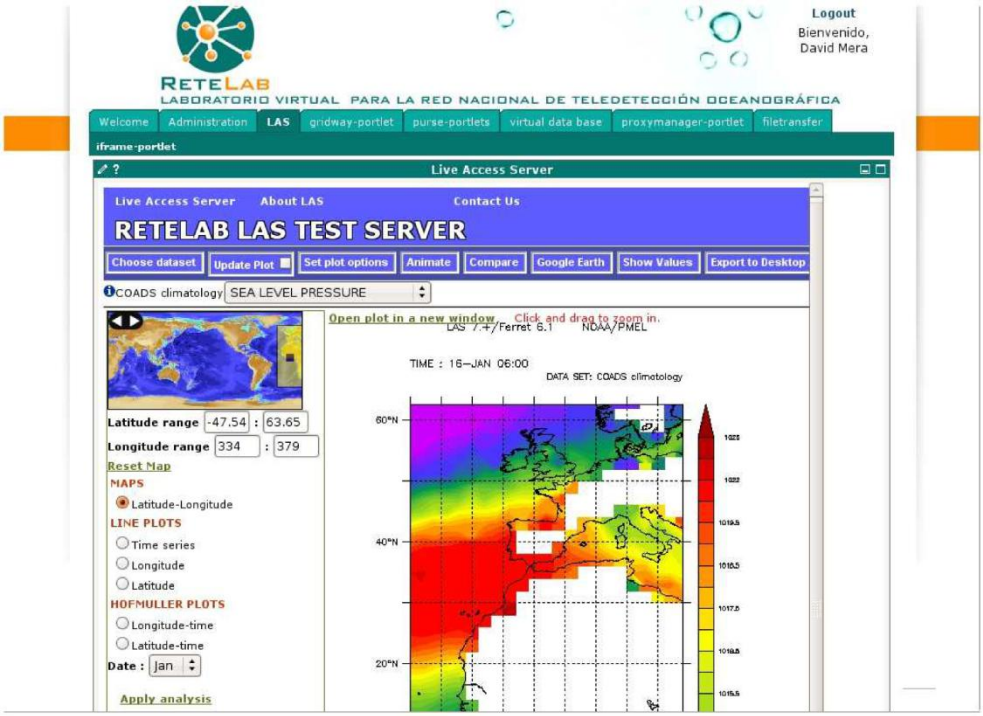
Retelab Distributed storage system

- Visualization tools



Integrated Data Viewer

Live Access Server

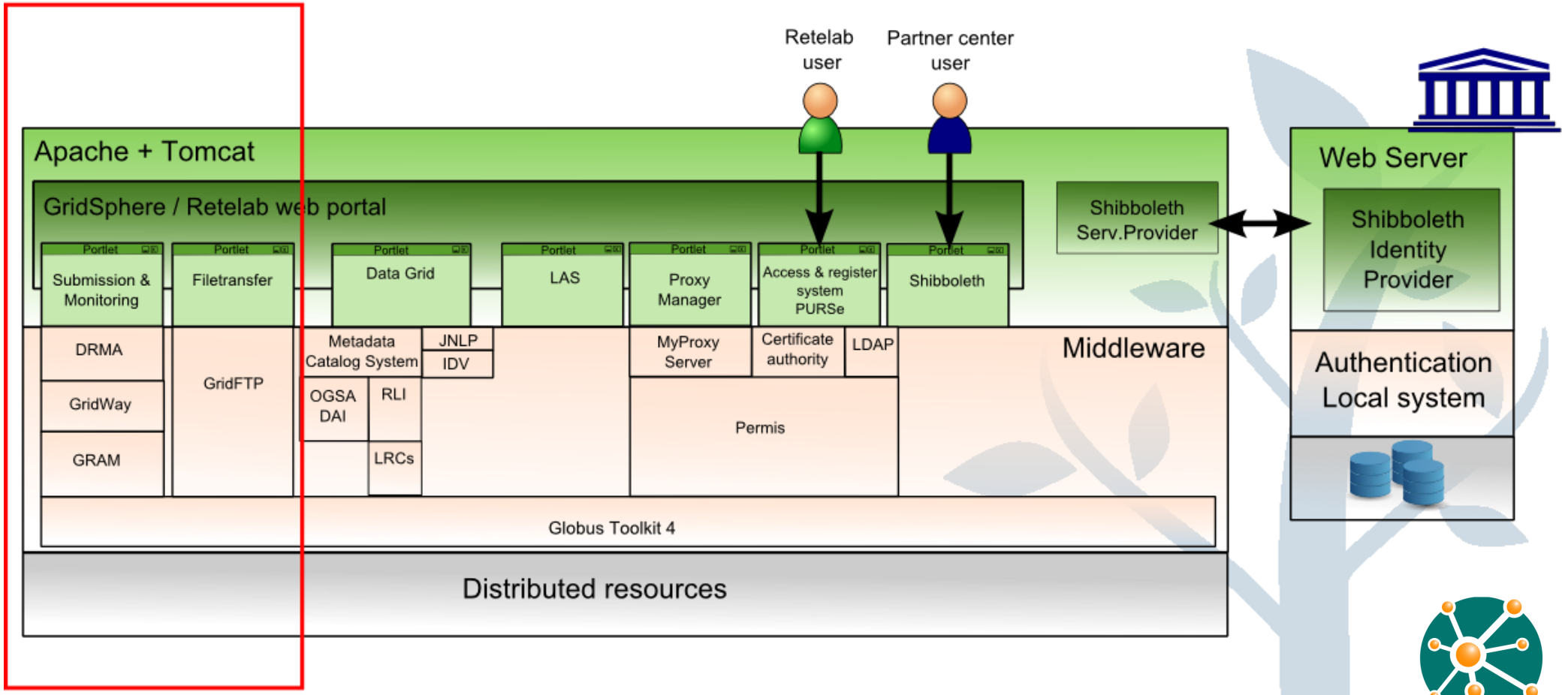


Retelab Job submission and monitoring system

- Traditional Grid systems
  - User interaction was required
  - Complex submission systems
- RETELAB
  - Grid metasqueduler
    - 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

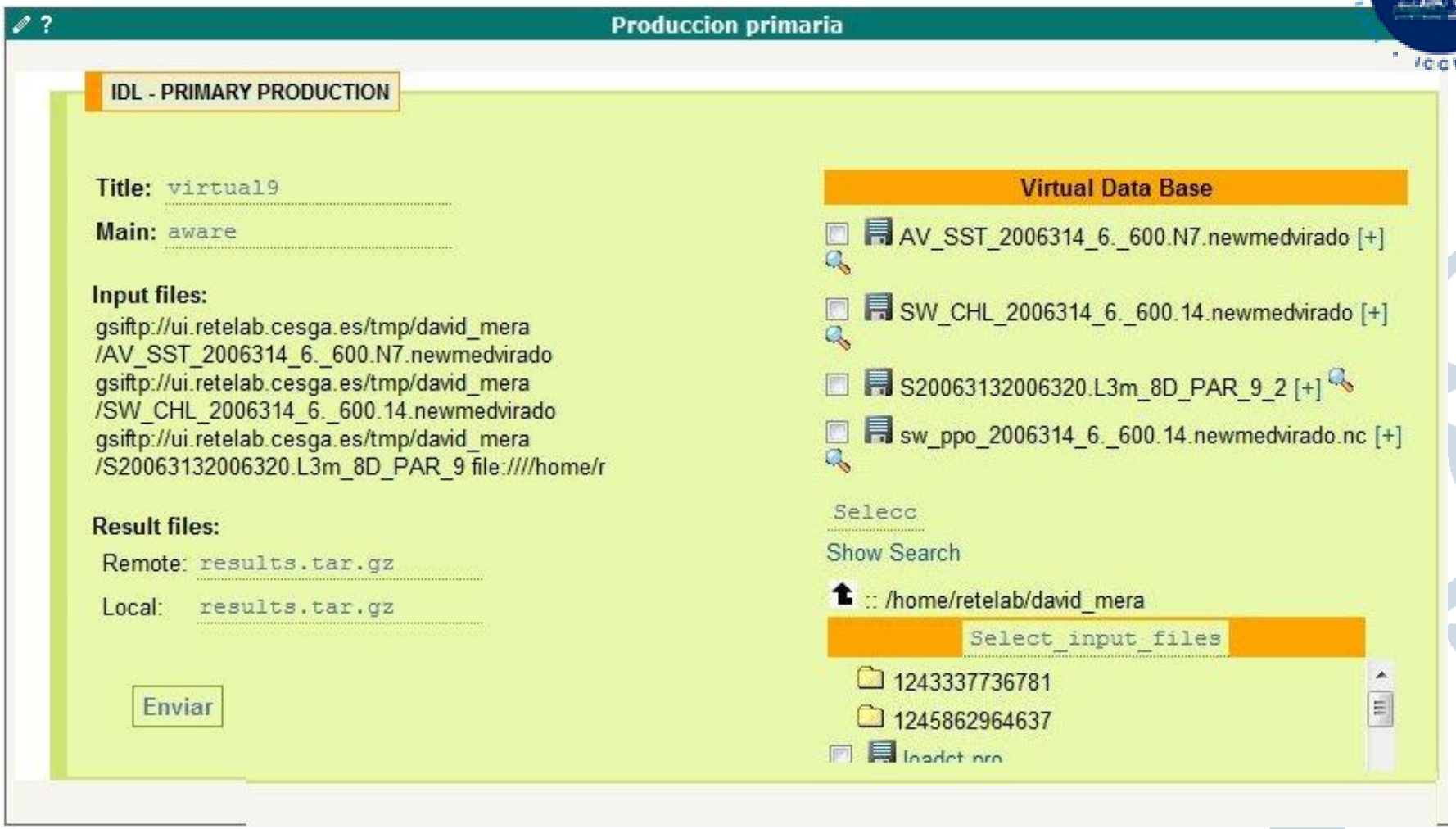
Job submission and monitoring system



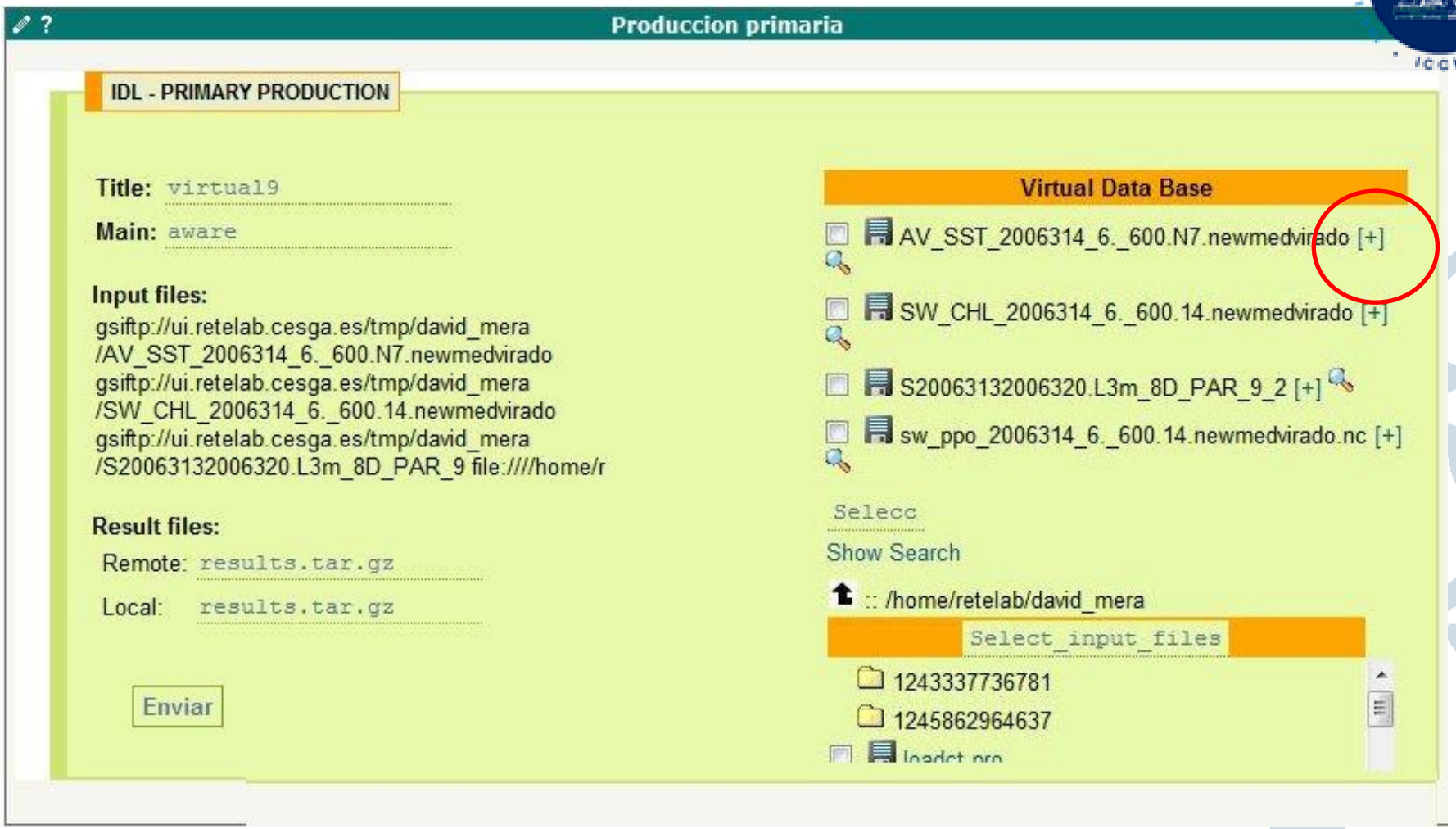
RETELAB

Retelab Use case

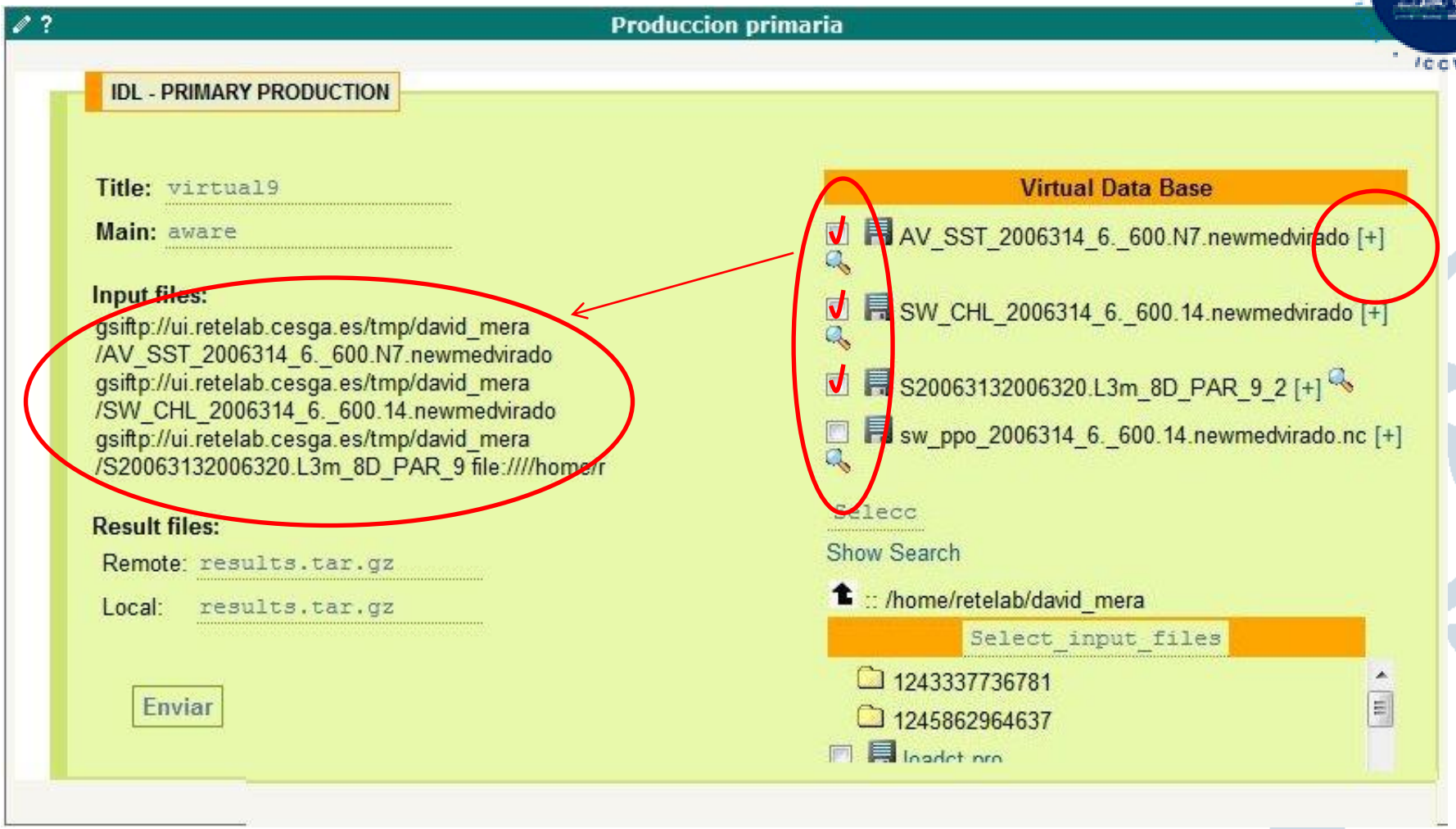
Retelab Use case



Retelab Use case

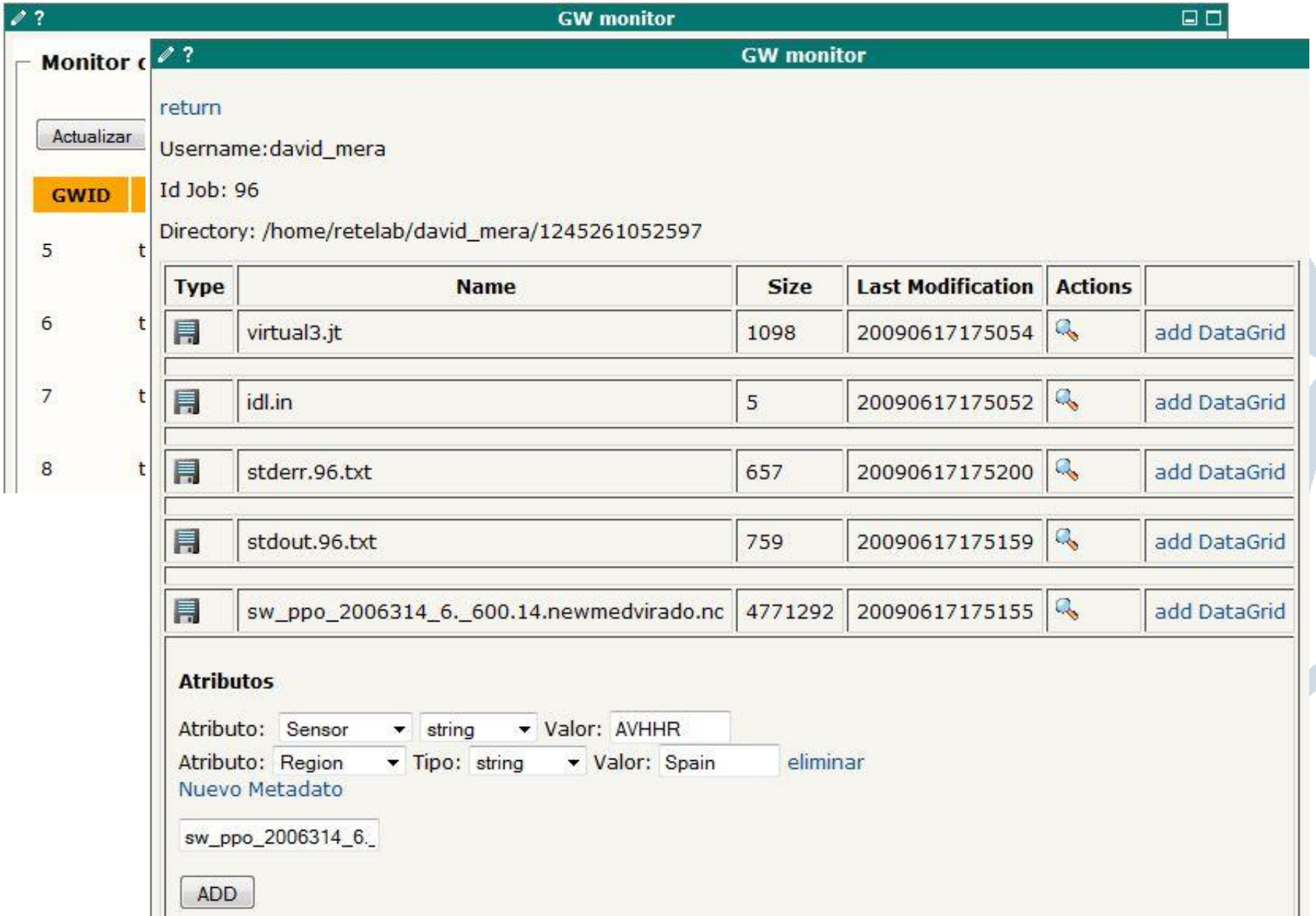


Retelab Use case

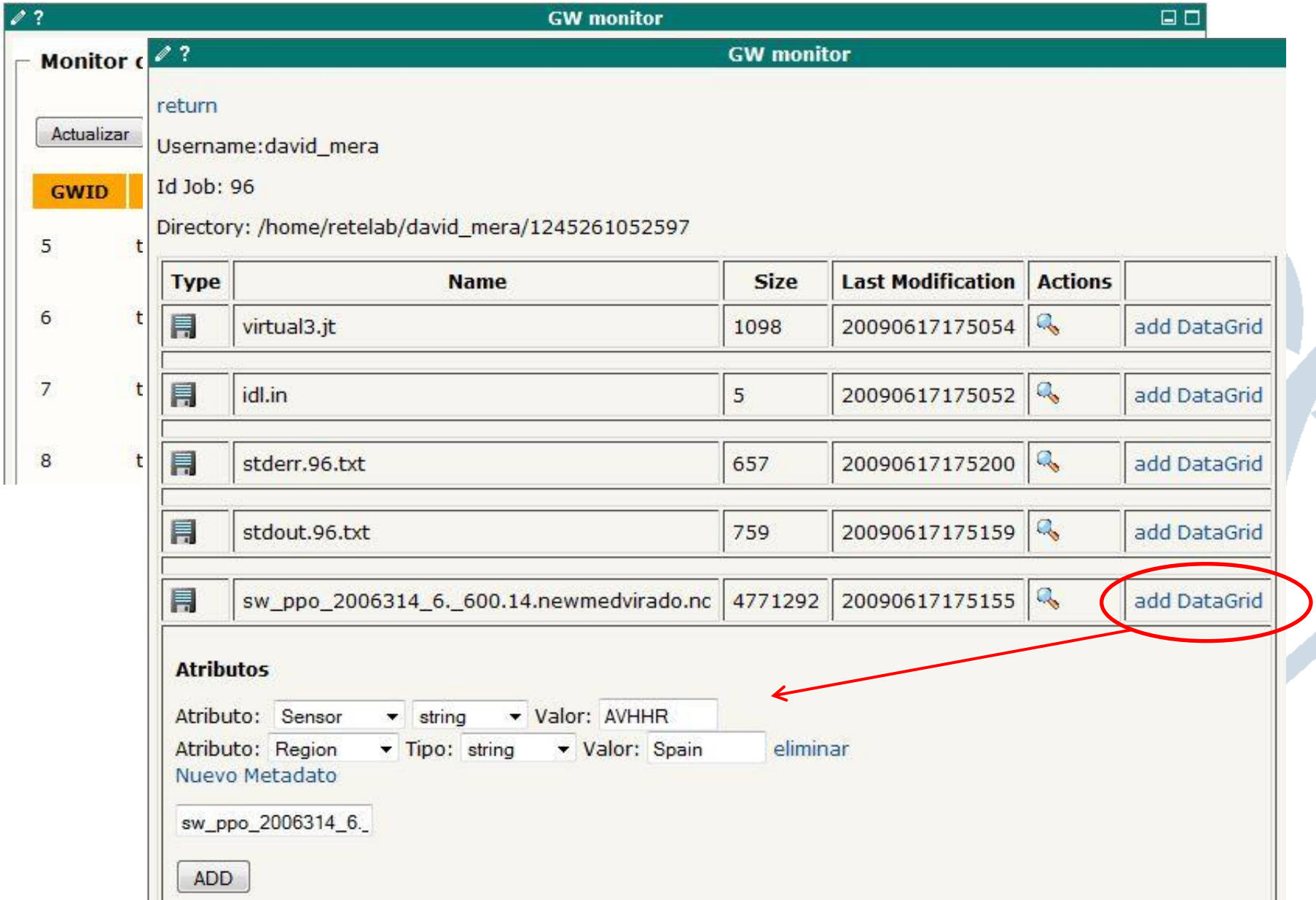


Retelab Use case

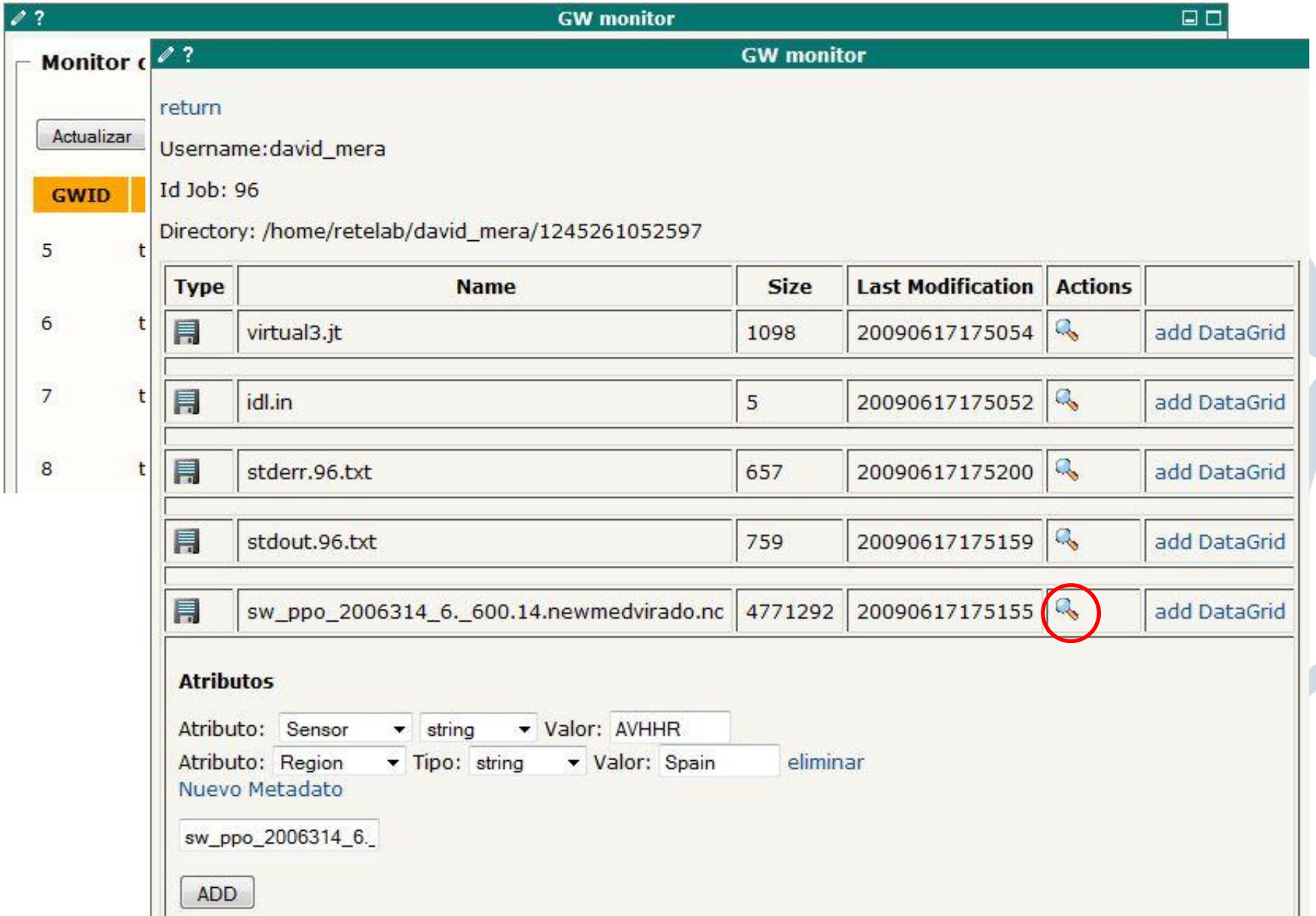
Retelab Use case



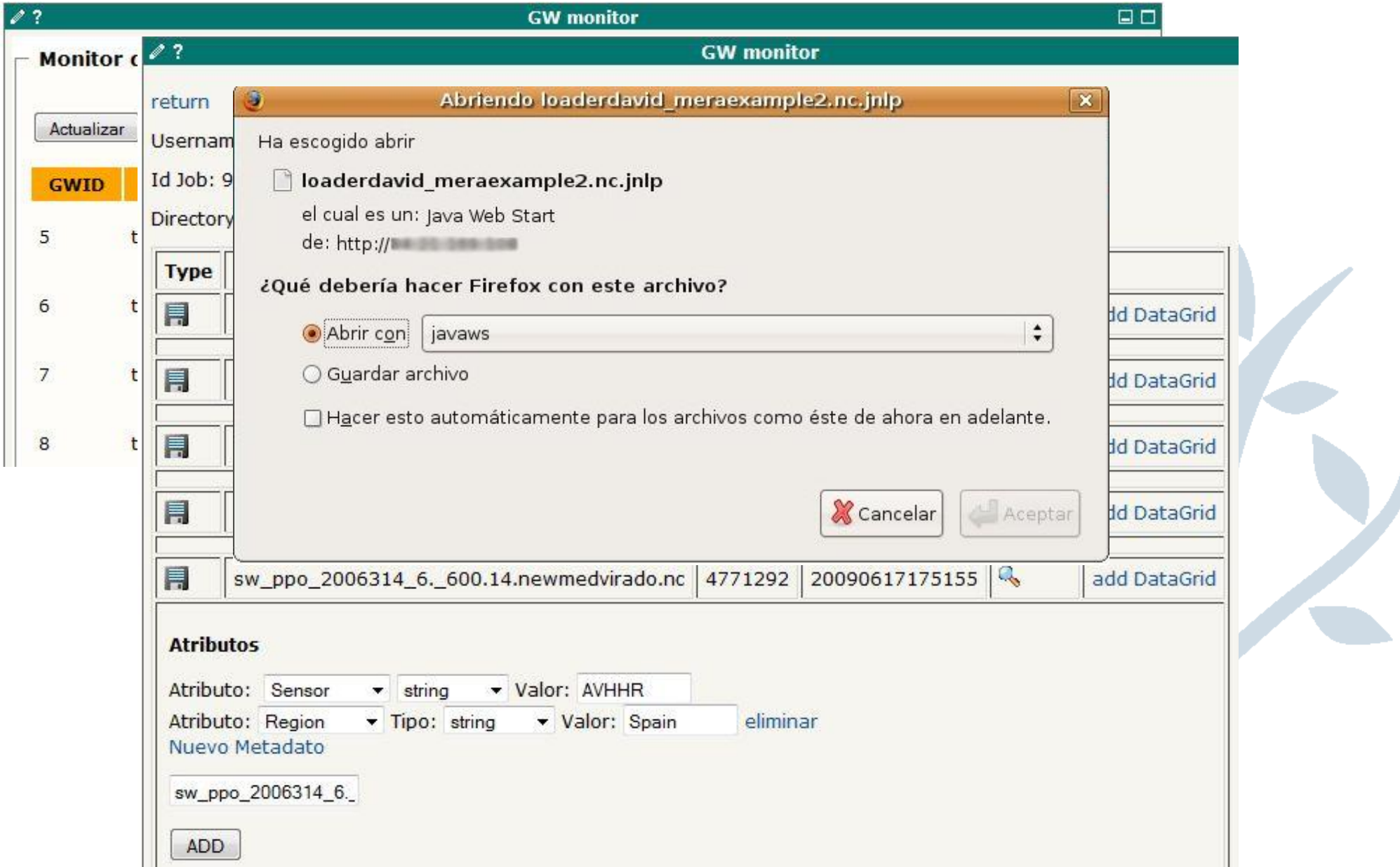
Retelab Use case



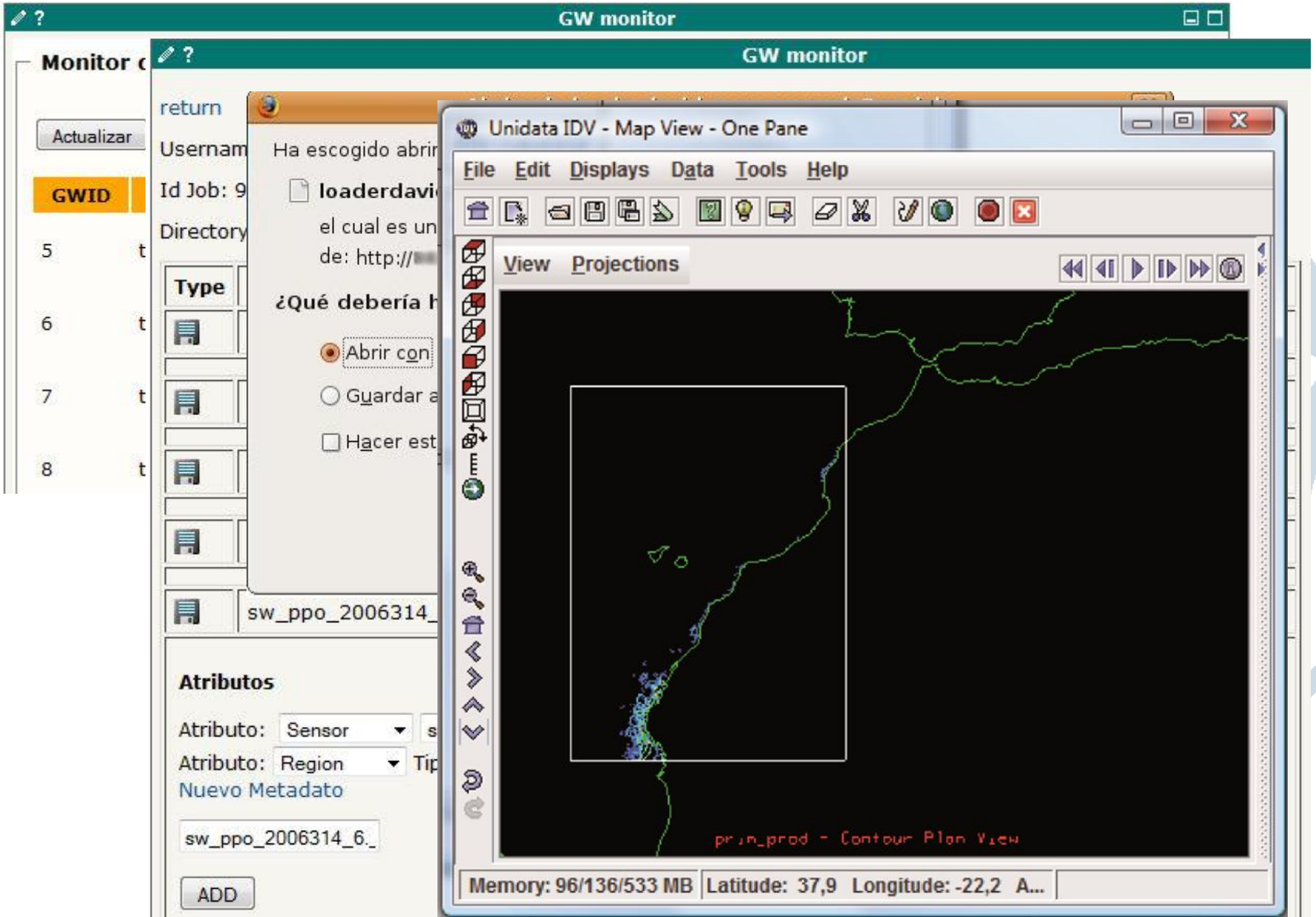
Retelab Use case



Retelab Use case



Retelab Use case

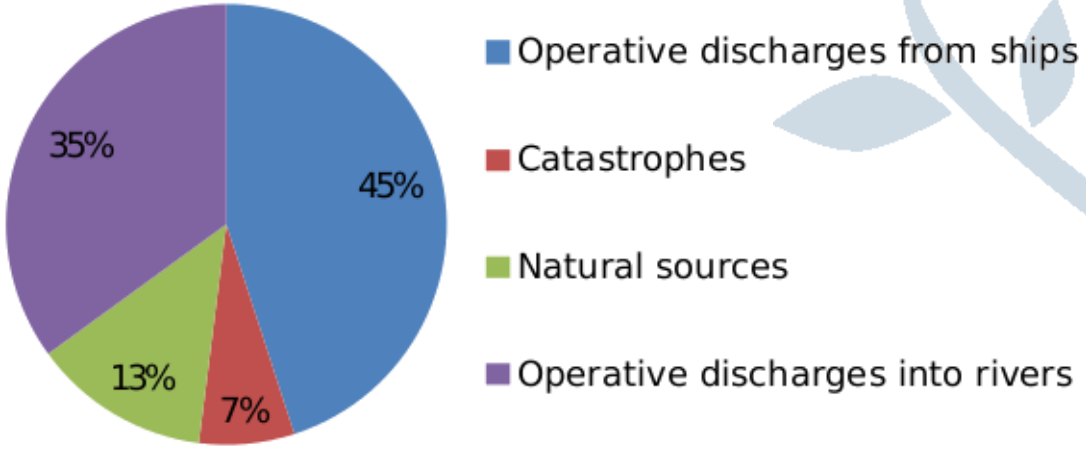


Index

Sentinazos

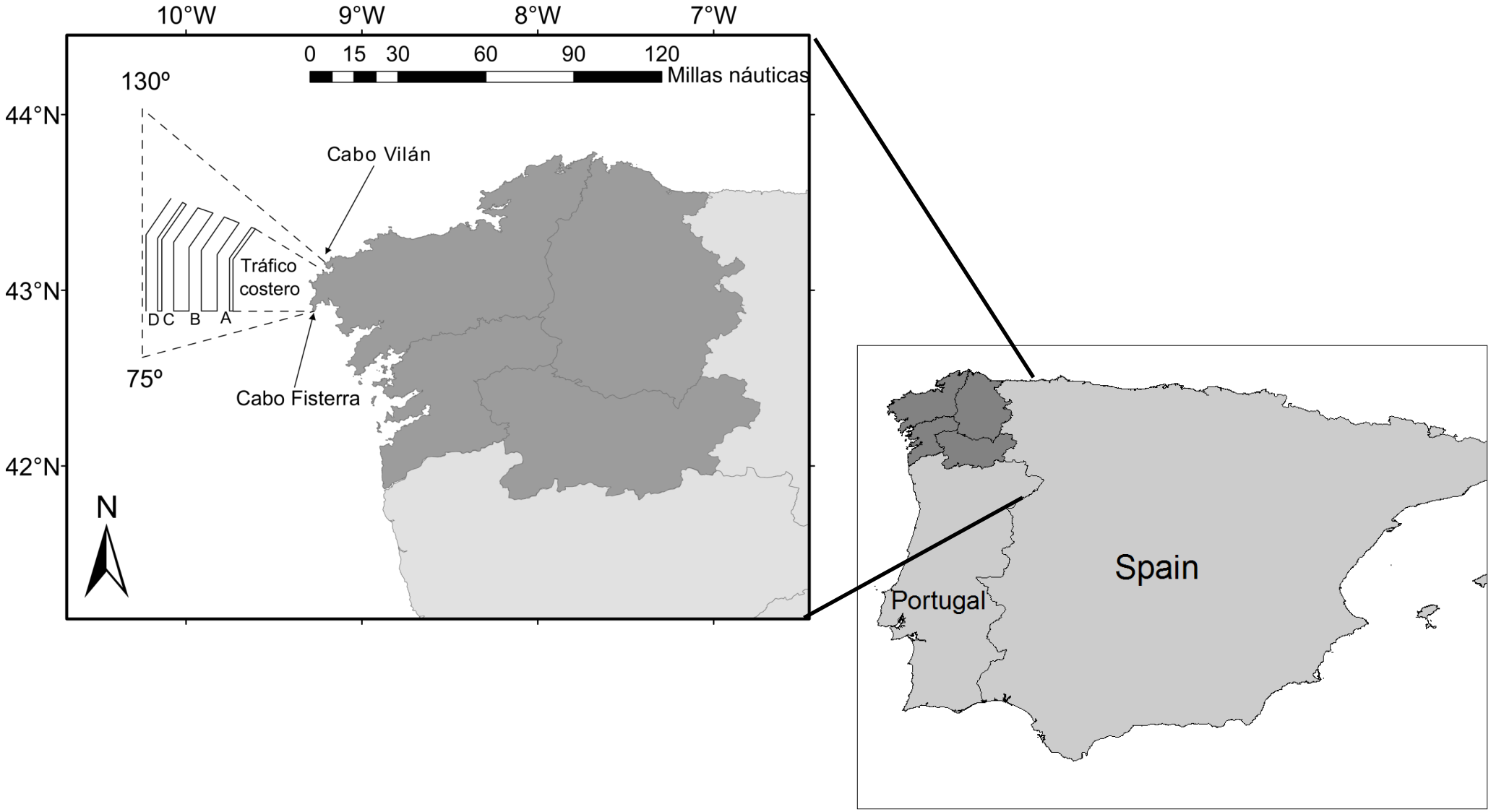
Introduction

- 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

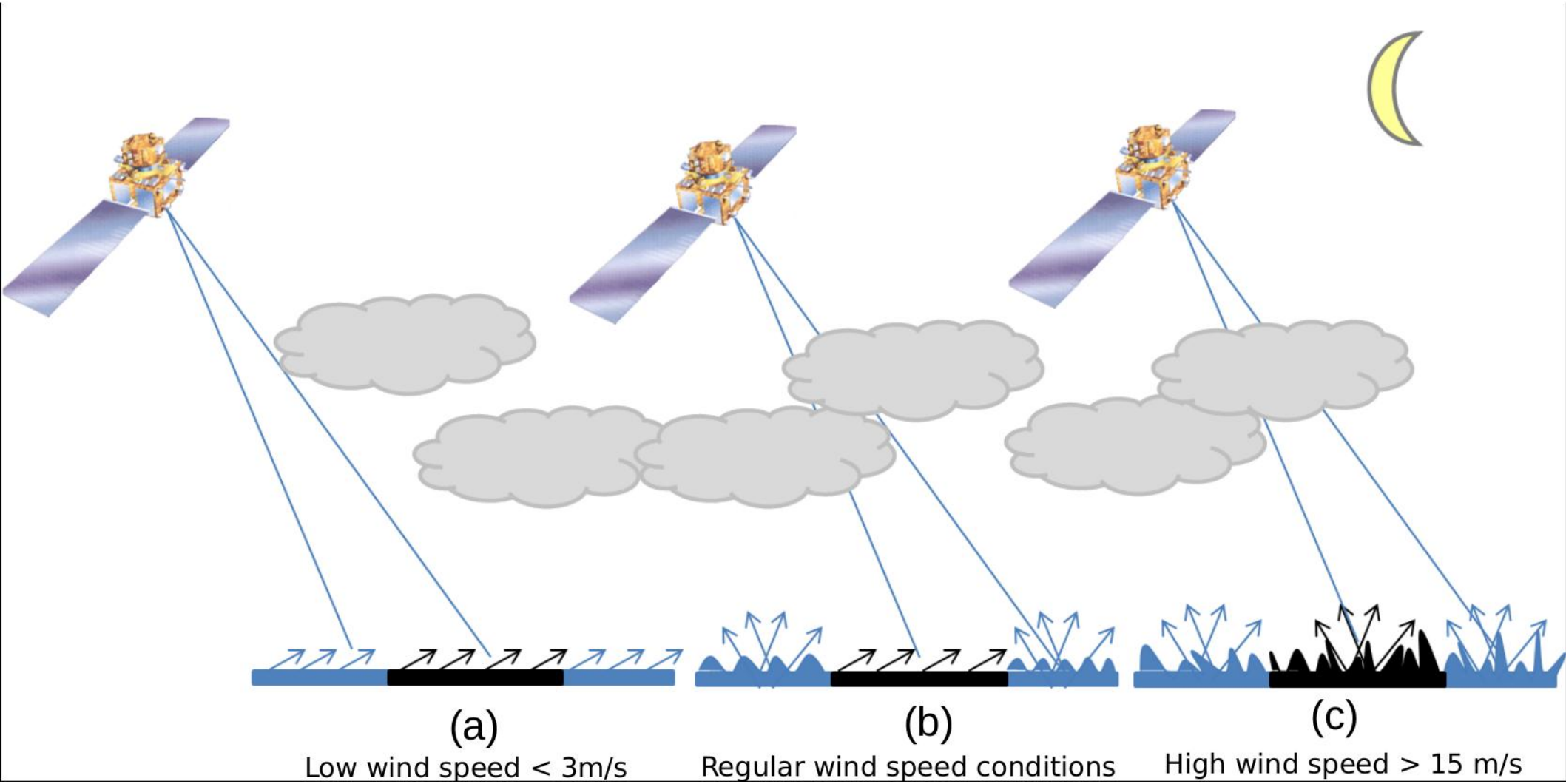
Introduction

Sentinazos

Introduction

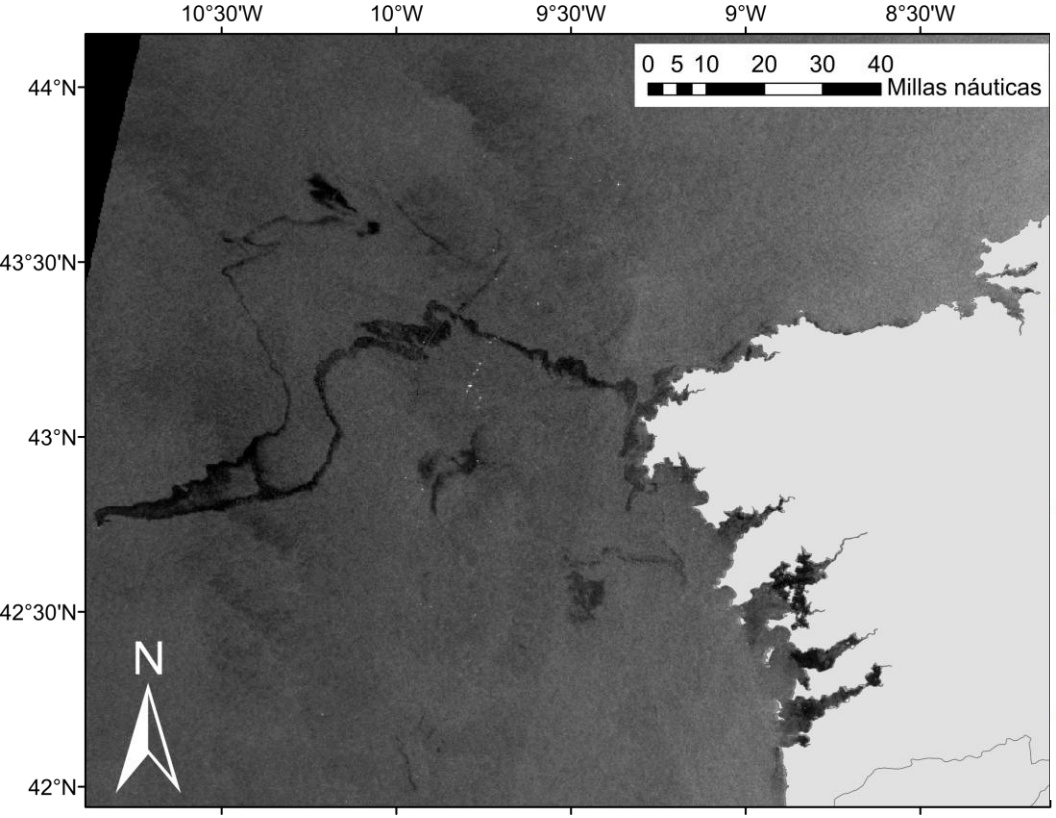
Sentinazos Introducción

- Synthetic Aperture Radar

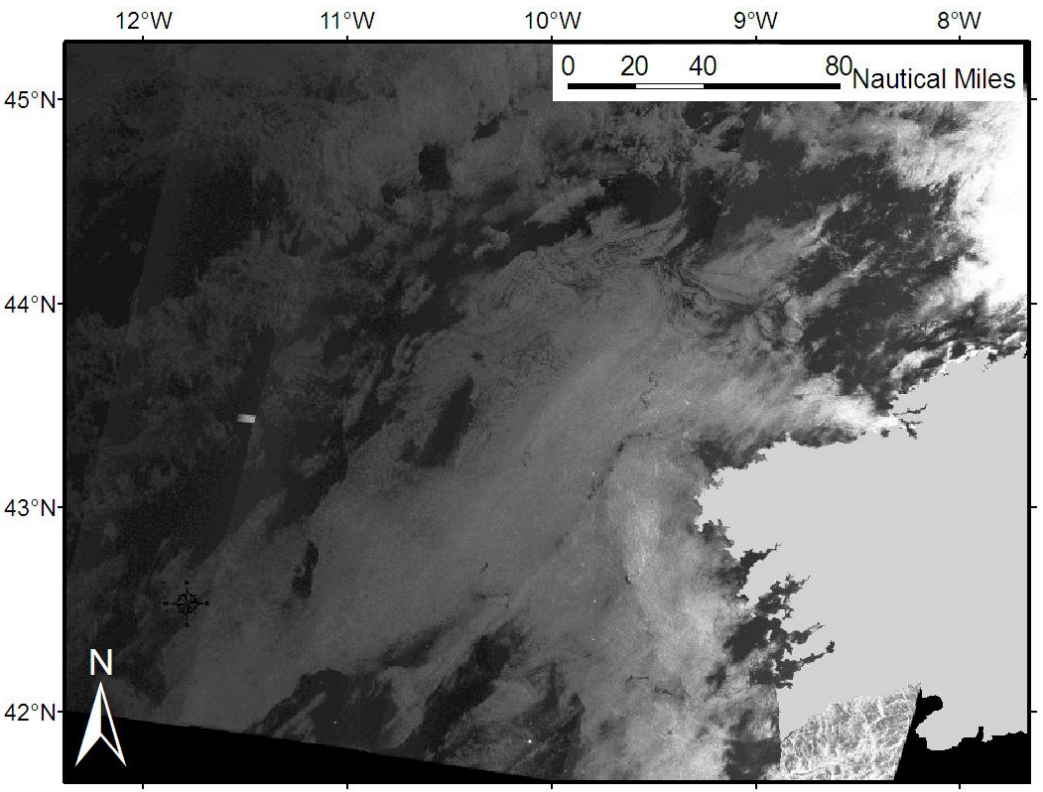


Sentinazos Introduction

- 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

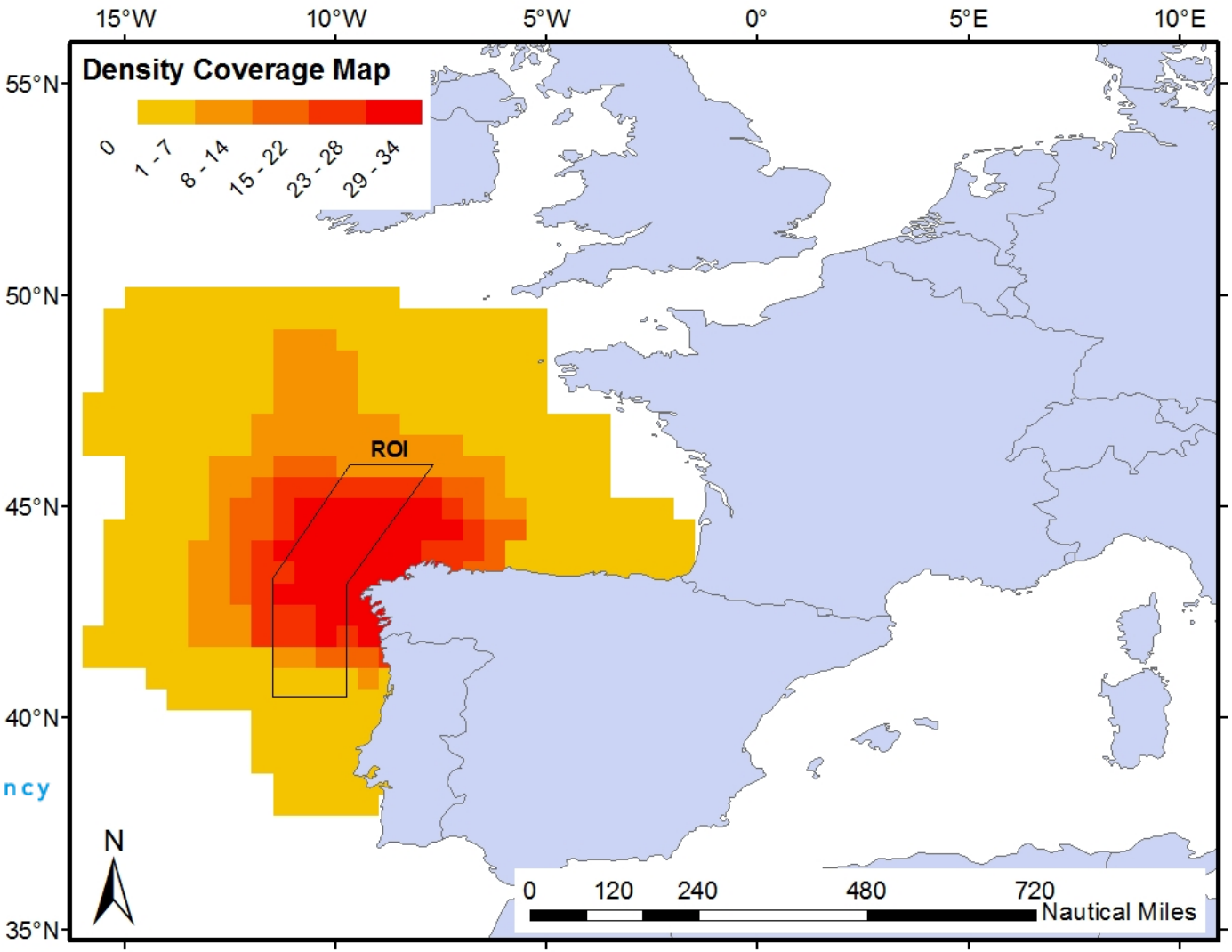
2. Oil spill shape could be relevant to correctly classify them

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



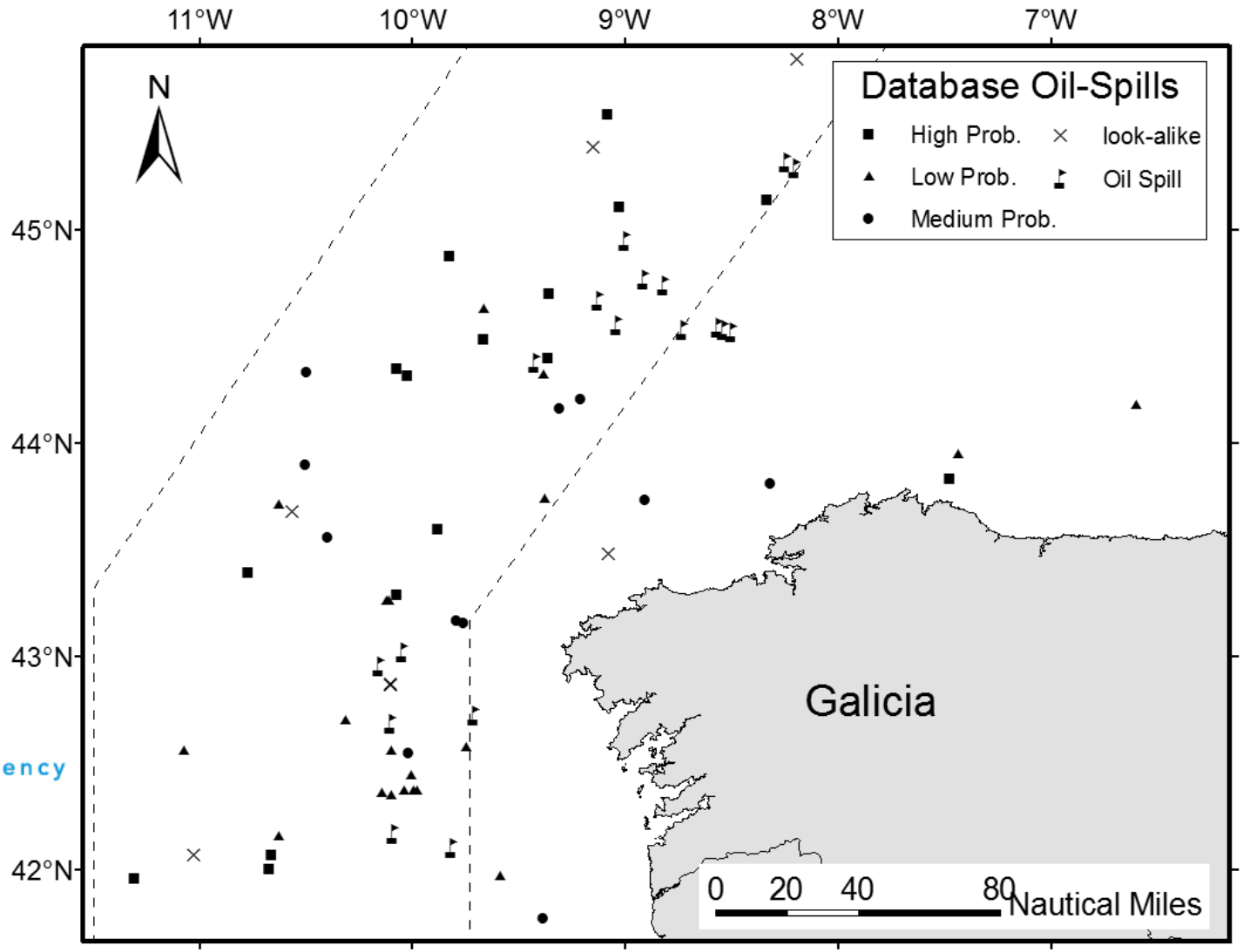
SALVAMENTO MARÍTIMO

esa

EMSA
European Maritime Safety Agency

Sentinazos Methodology

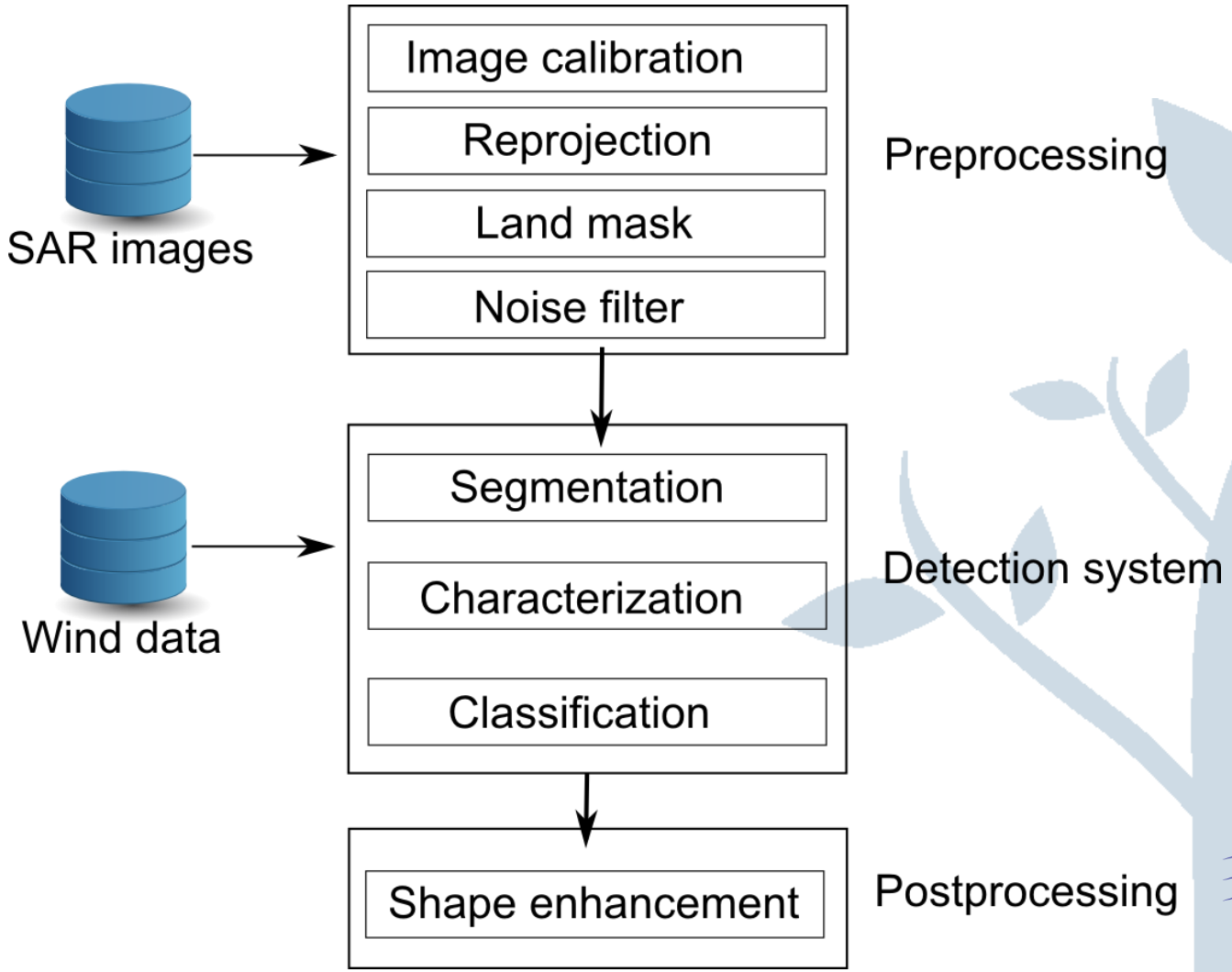
- Dataset- 47 SAR images from the Envisat



European Maritime Safety Agency

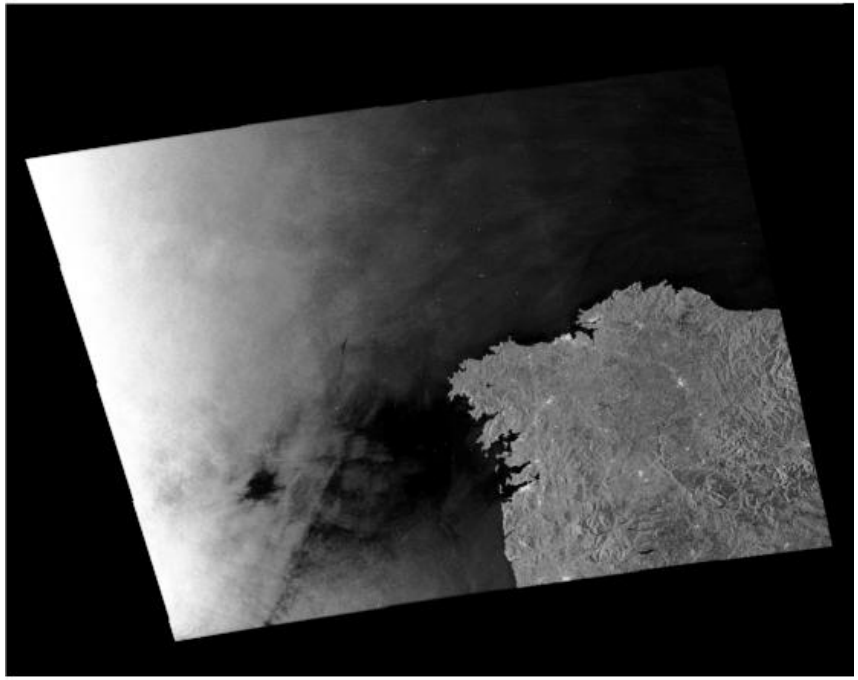
Sentinazos Methodology

- System architecture

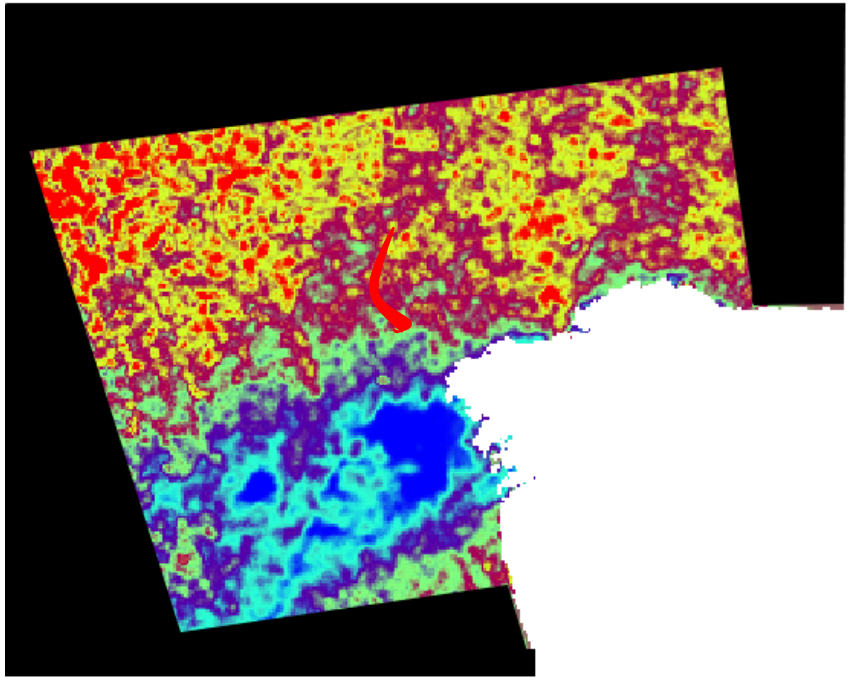


Sentinazos

Methodology - Segmentation



SAR image. Envisat, 01/06/2007

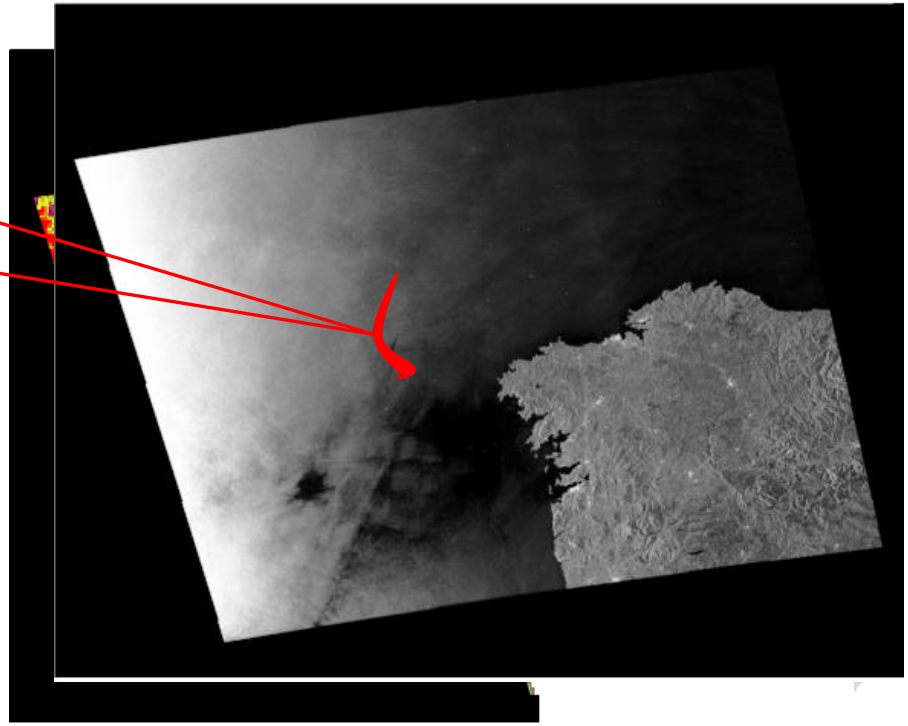


Wind intensity data. CMOD5 model

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



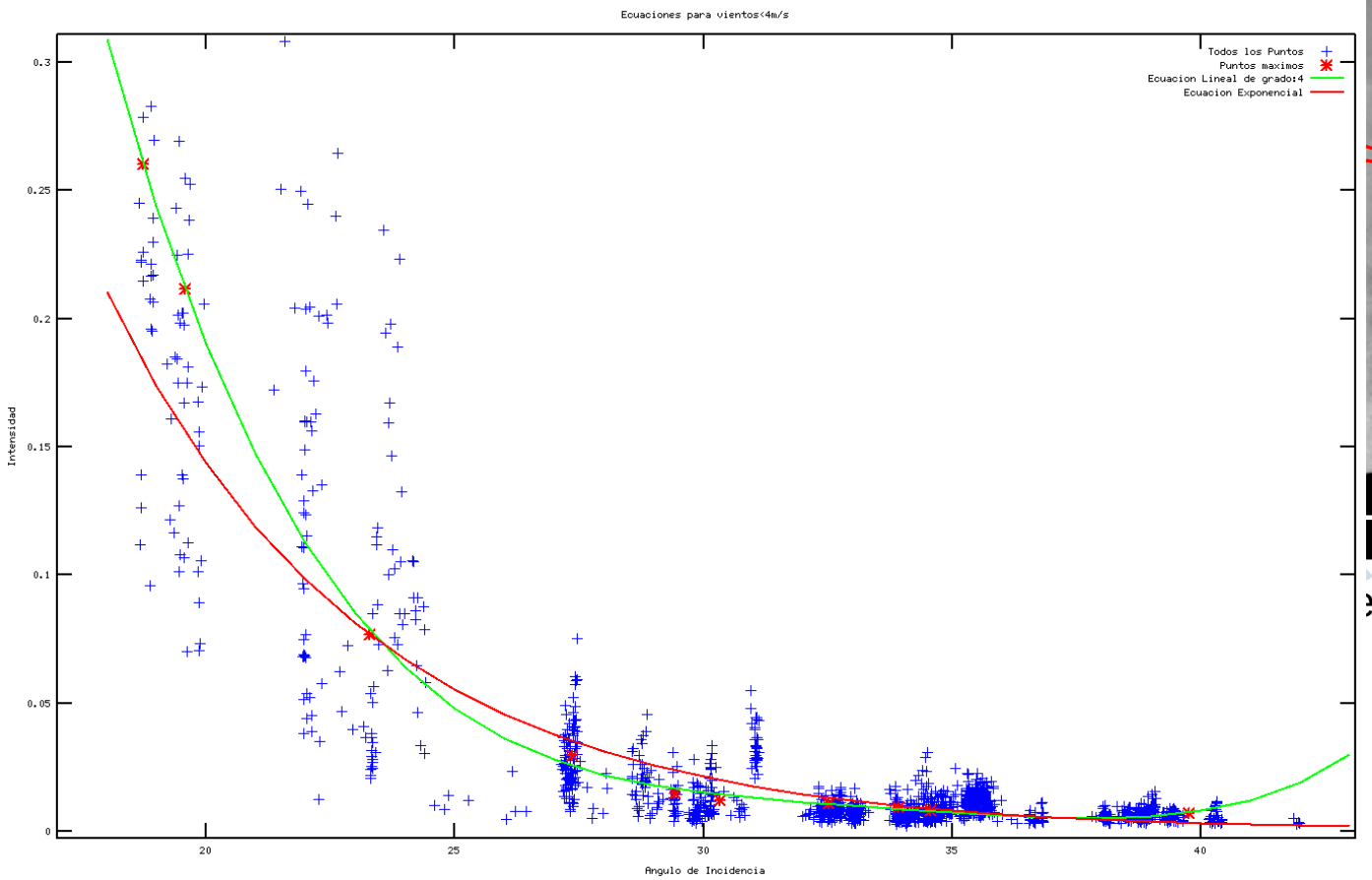
SAR image. Envisat, 01/06/2007

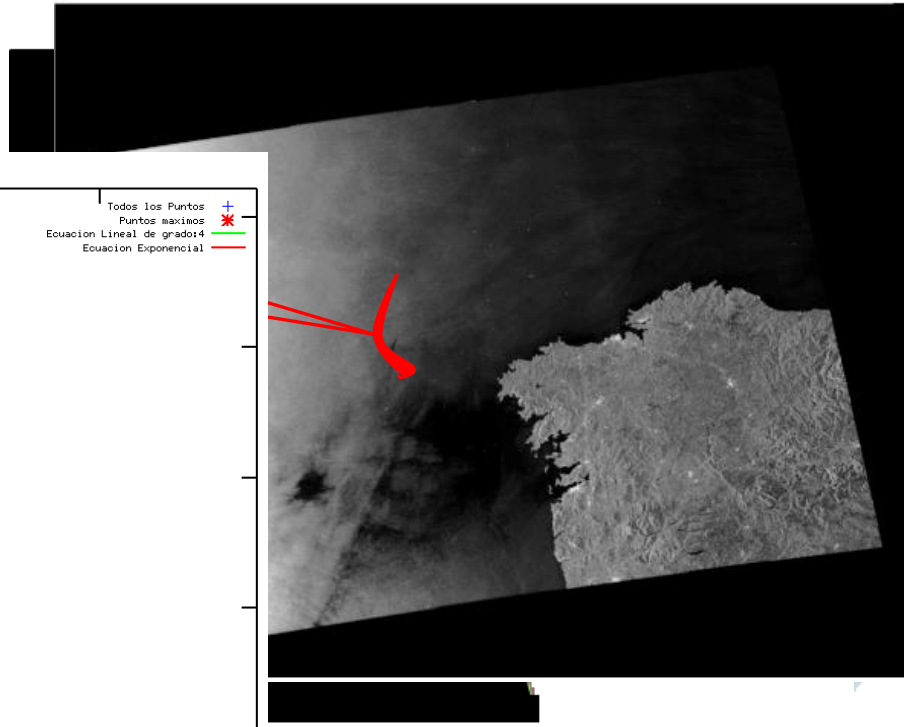
Wind intensity data. CMOD5 model

Sentinazos

Methodology - Segmentation

X	Y	Wind	I.A.	Intensity



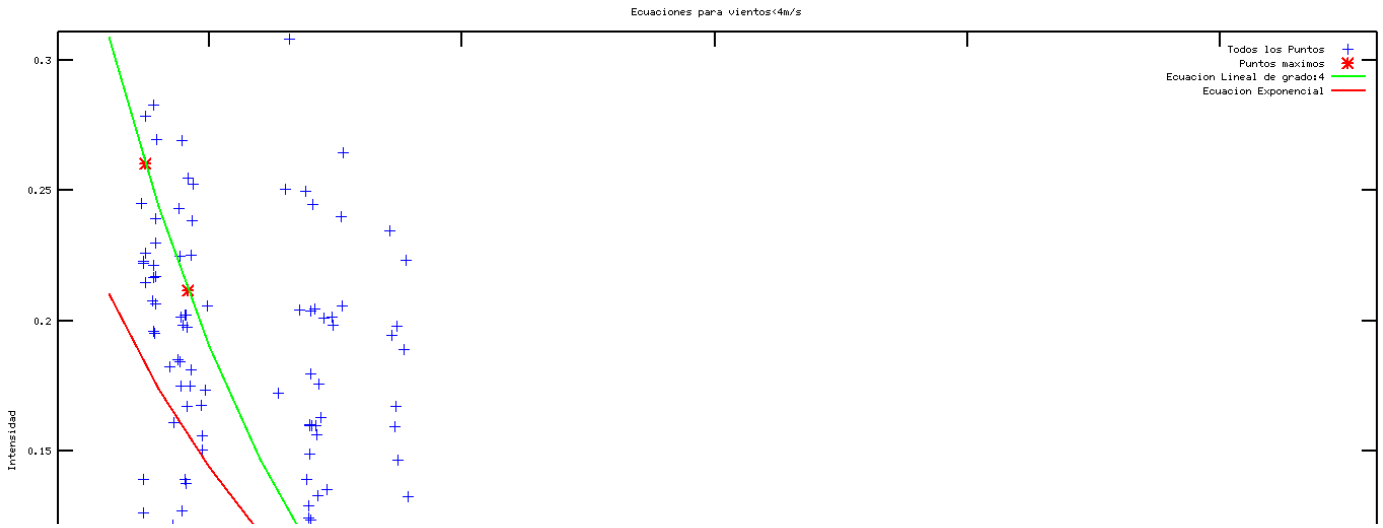


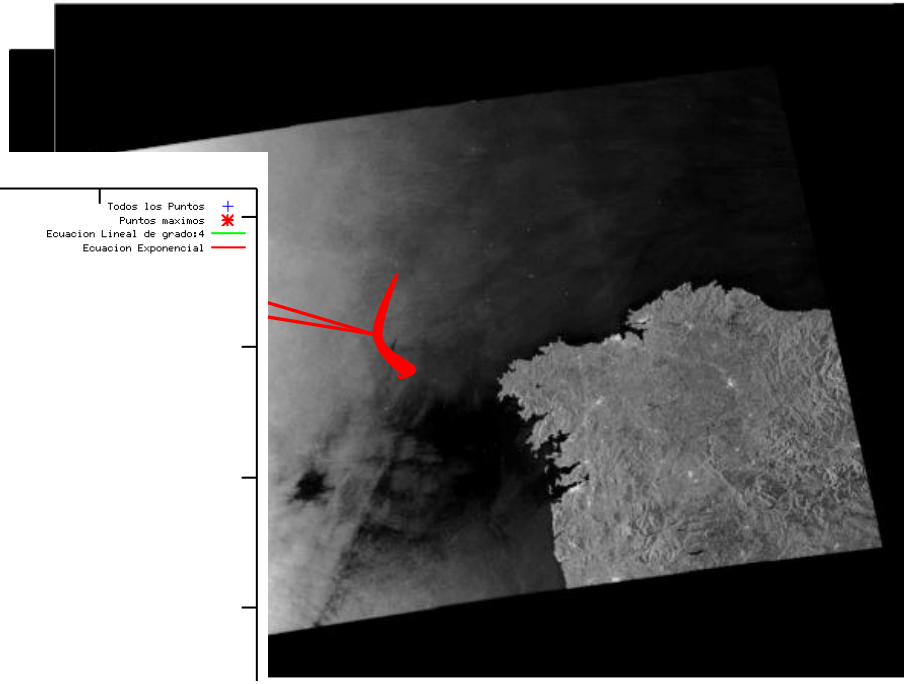
ensity data. CMOD5 model

Sentinazos

Methodology - Segmentation

X	Y	Wind	I.A.	Intensity

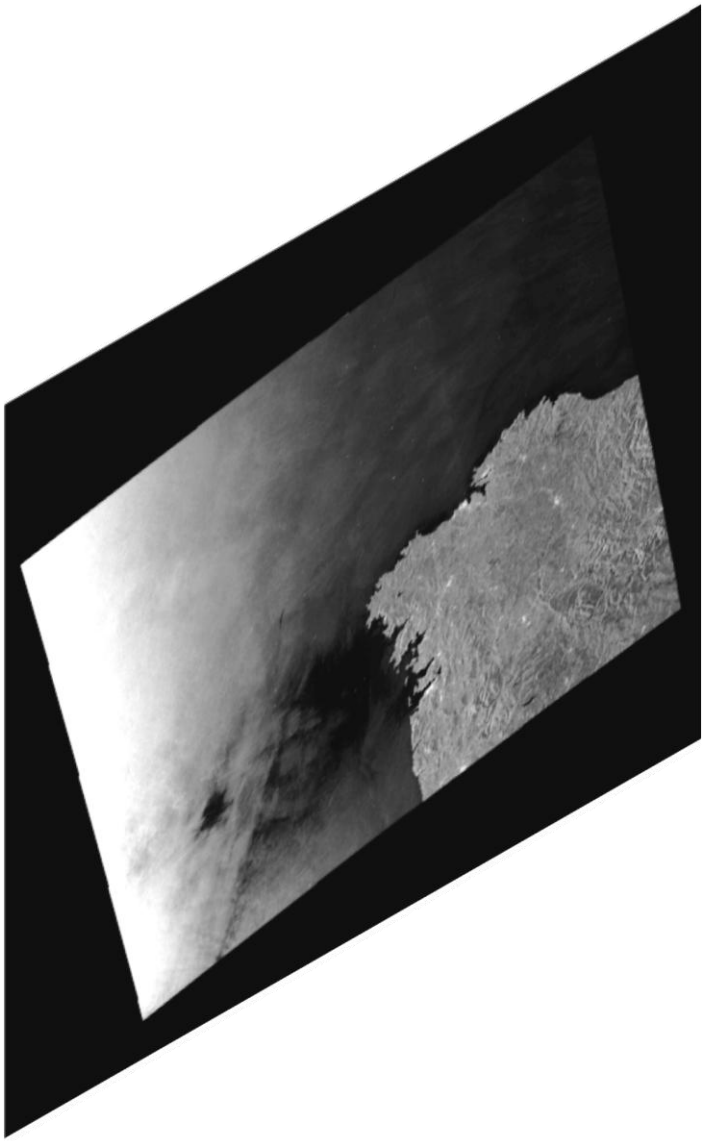




Wind speed	Quartic function	Cutoff point	Negative exponential function
$vel. <= 4m/s$	$6,3662 * 10^{-6}x^4 - 0,00083671x^3 + 0,041262x^2 - 0,90719x + 7,5353$	$37,8°$	$5,9169e^{-0,17881x}$
$vel. <= 5m/s$	$3,0874 * 10^{-6}x^4 - 0,00043845x^3 + 0,023424x^2 - 0,55931x + 5,0588$	$36,25°$	$6,6689e^{-0,17944x}$
$vel. <= 6m/s$	$2,1022 * 10^{-6}x^4 - 0,00031131x^3 + 0,017414x^2 - 0,4365x + 4,1503$	$36,82°$	$5,7770e^{-0,17217x}$
$vel. <= 7m/s$	$2,7742 * 10^{-6}x^4 - 0,00039094x^3 + 0,02086x^2 - 0,50088x + 4,5883$	$37,52°$	$5,4477e^{-0,16858x}$
$vel. > 7m/s$	$1,2416 * 10^{-6}x^4 - 0,00019466x^3 + 0,011604x^2 - 0,31143x + 3,182$	$36,8°$	$6,2053e^{-0,17101x}$

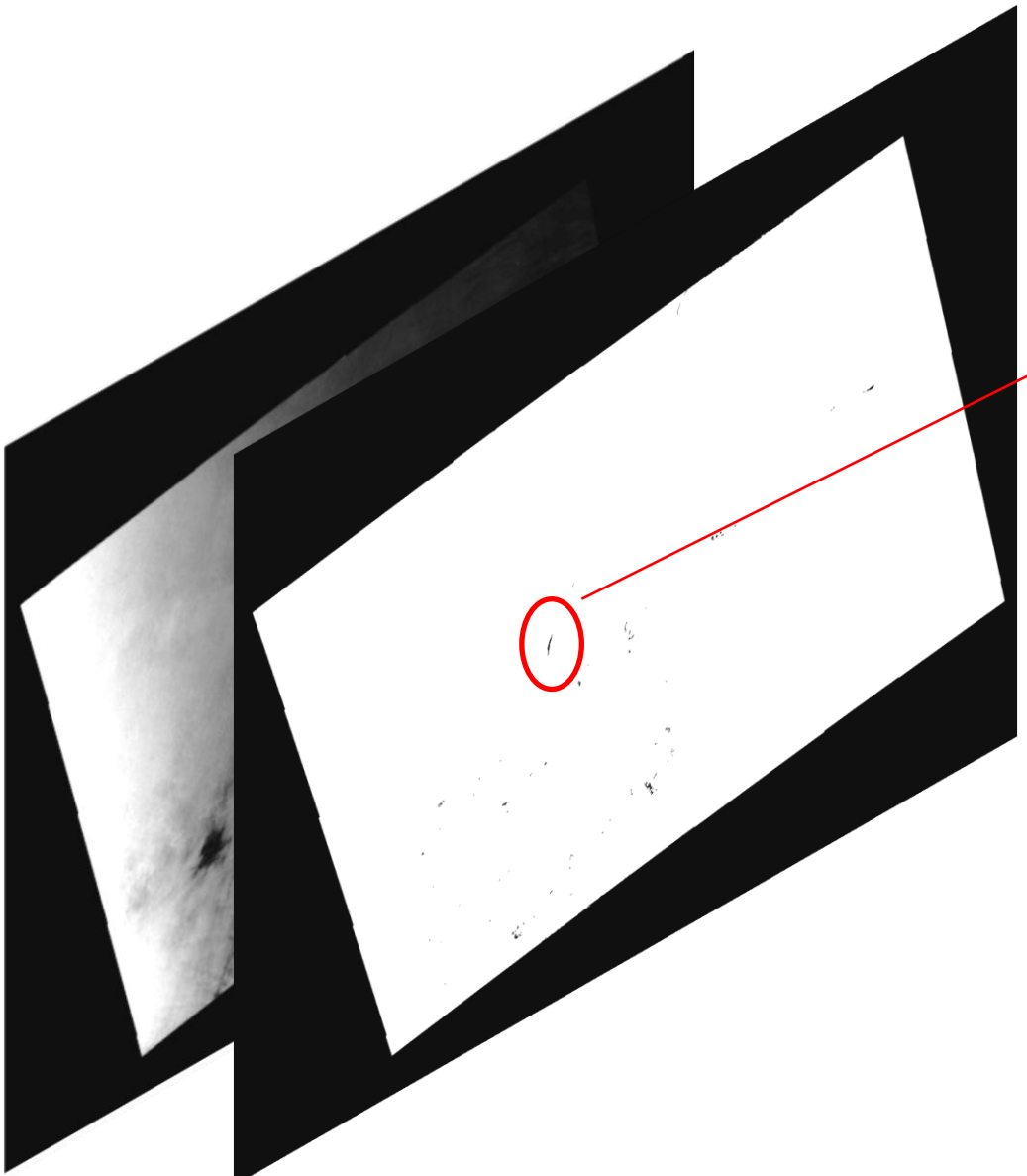
Sentinazos

Methodology - Characterization

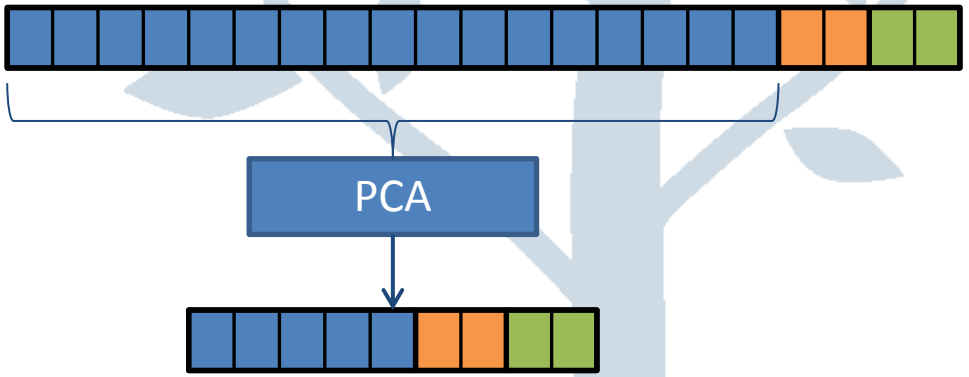


Sentinazos

Methodology - Characterization



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



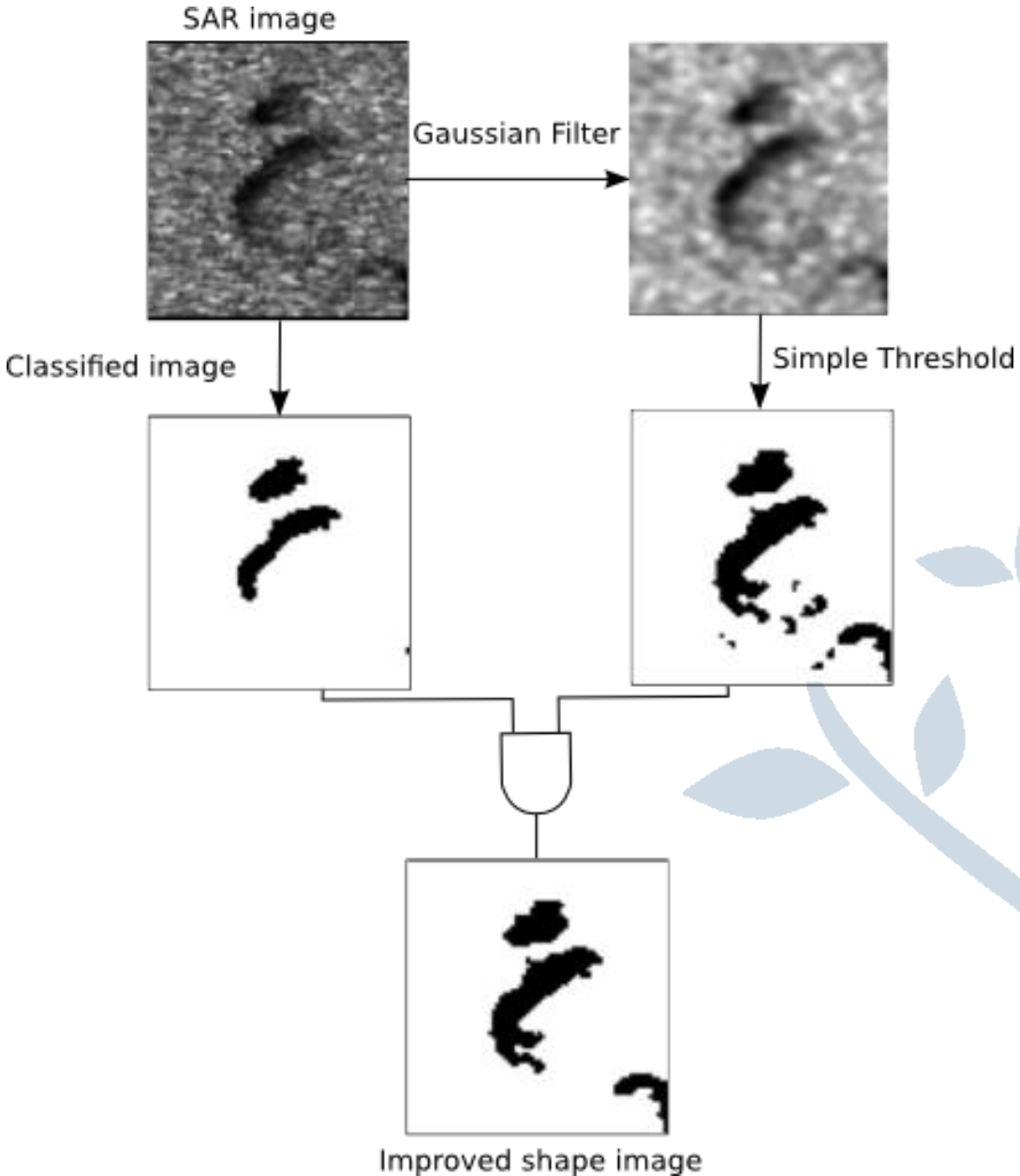
Sentinazos

Methodology - Classification

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

Sentinazos

Methodology - Postprocessing



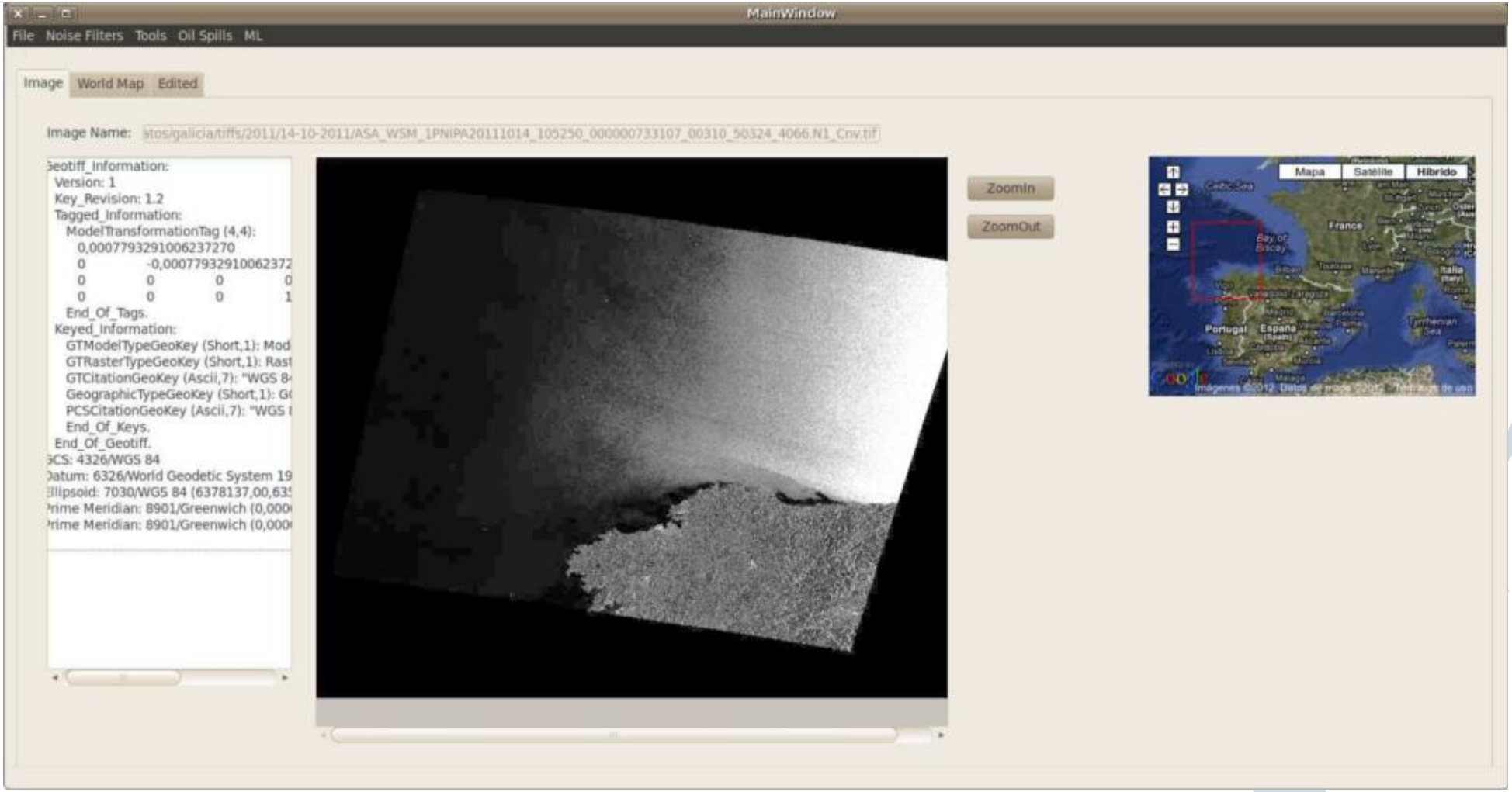
Sentinazos Results

	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%

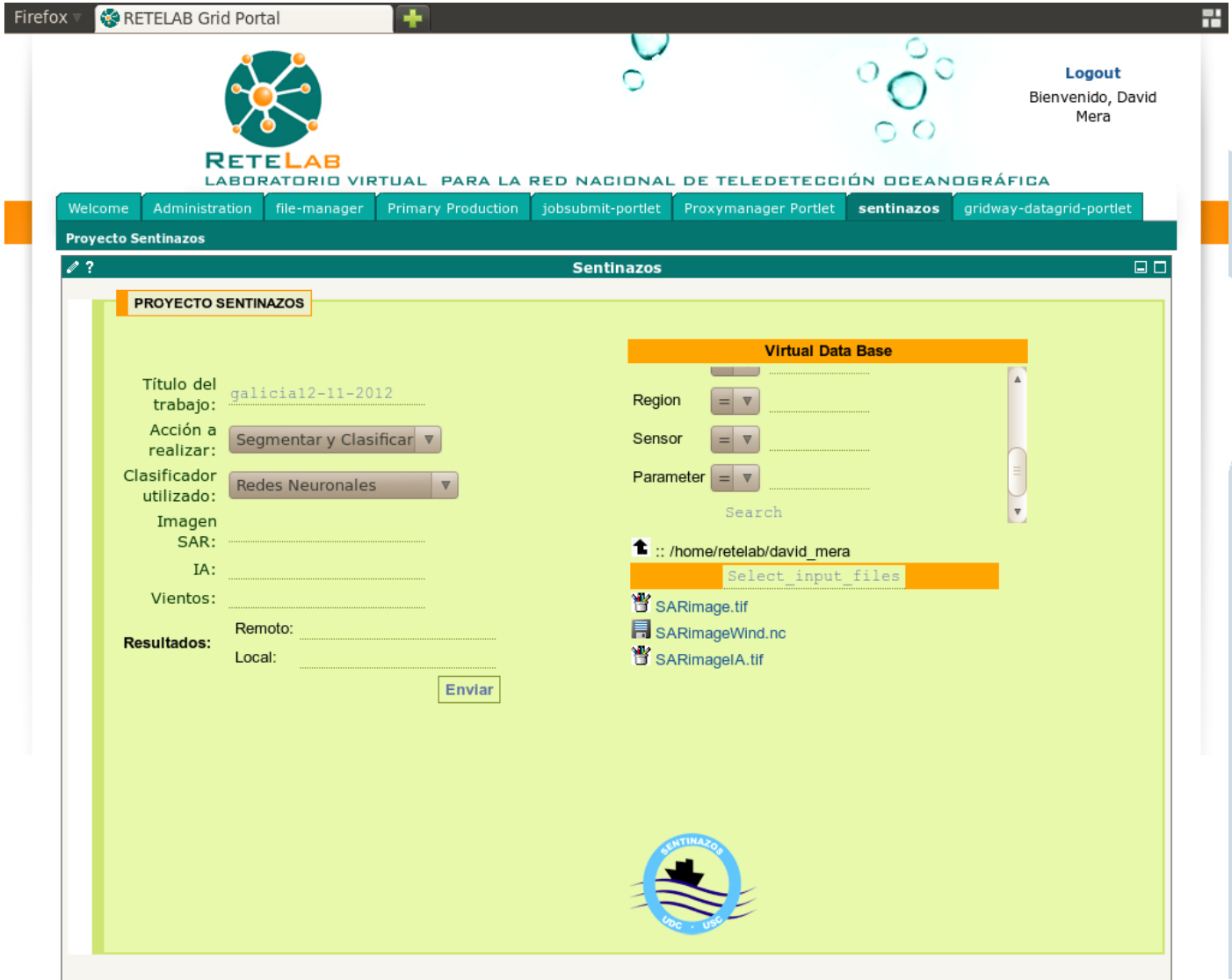
Sentinazos Results

- Desktop application

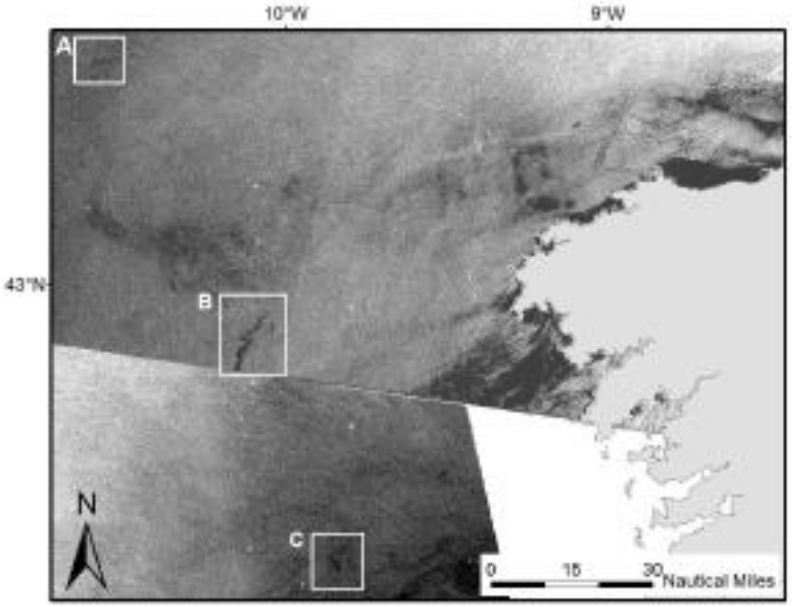


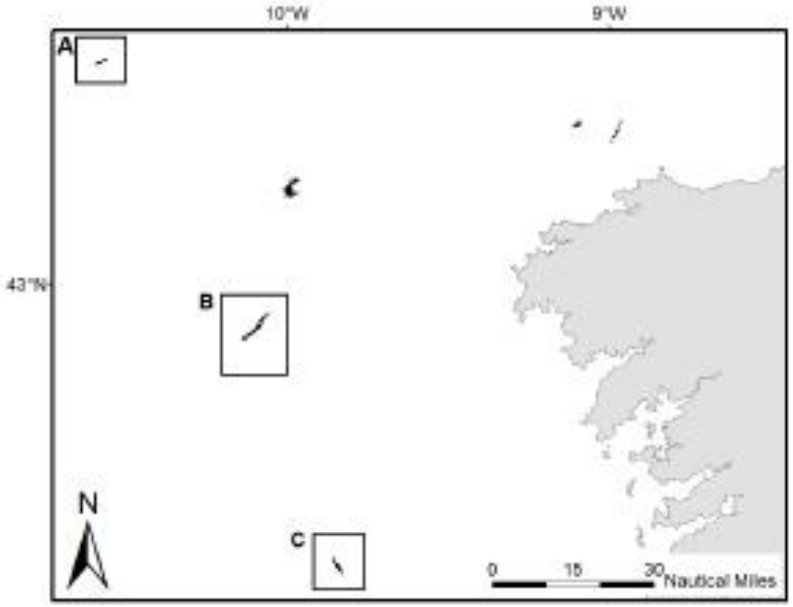
Sentinazos Results

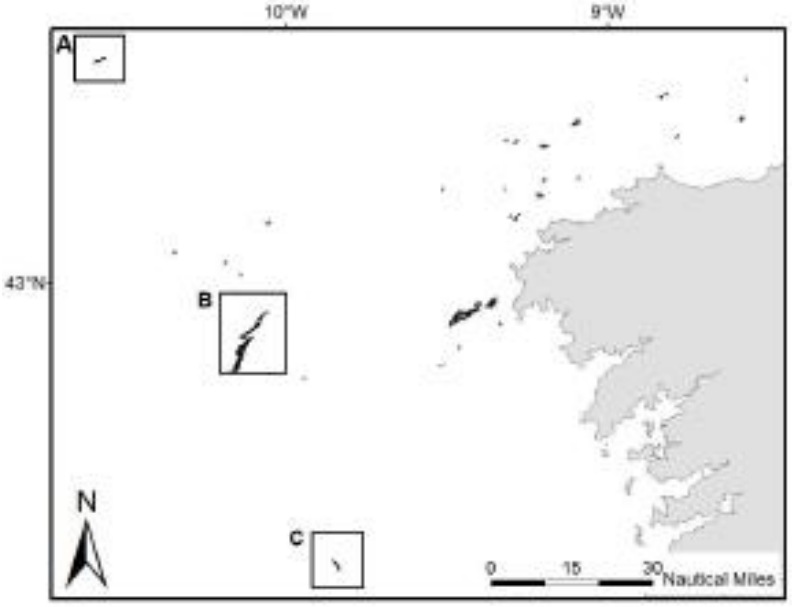
- Portlet Grid

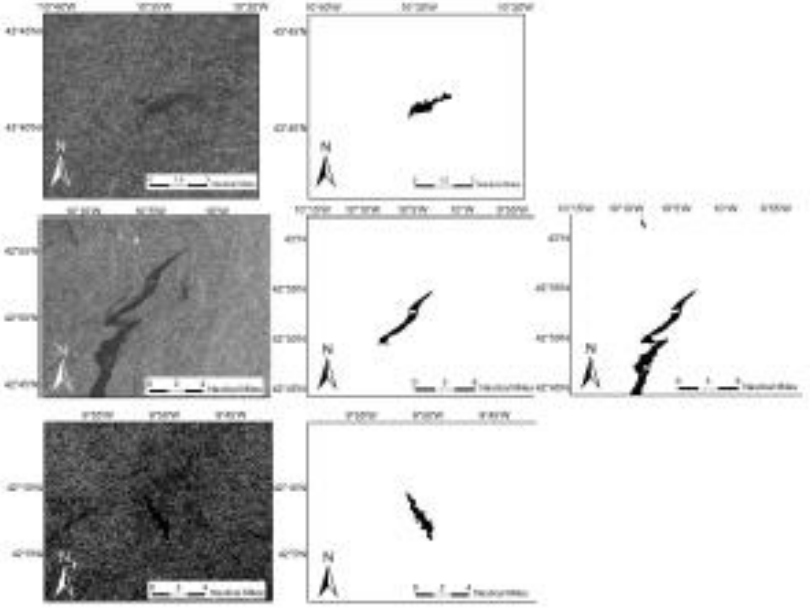


Sentinazos Results









Future

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

Thank you

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

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Instituto de Investigacións Tecnolóxicas

CiTiUS Centro Singular de Investigación en Tecnoloxías da Información