Online data mining services for dynamic spatial databases

Manuel Costa, Inês Sousa, Alexandra Fonseca, Luís Teixeira, Alvaro Barbosa, Jorge Cardoso, Vasco Carvalho, Ivan Franco, Nuno Capeta, Marco Orlando, Paulo Castro, Diana Henriques and Paulo Rosa

Motivation

Large **databases updated by automatic processes**.

There aren't many commercial solutions to execute **data mining algorithms at server side**.

Automatic processes to analyze dynamic data represent an appealing market (e.g. **traffic**, **air quality**, **water quality**).

Goals

Develop a **Service Oriented Architecture** (SOA) with data mining functionalities

Direct access to dynamic spatial Data Bases and Data Mining algorithms

Develop customized applications

Outline

Brief overview of Data Driven Modeling

Data mining services in action : SNIRH Data Mining

Technical description of the services

Case Studies : Spatial sound data mining

Data Driven Modeling

Robust methods to model relations between data (e.g. Artificial Neural Networks ANN).

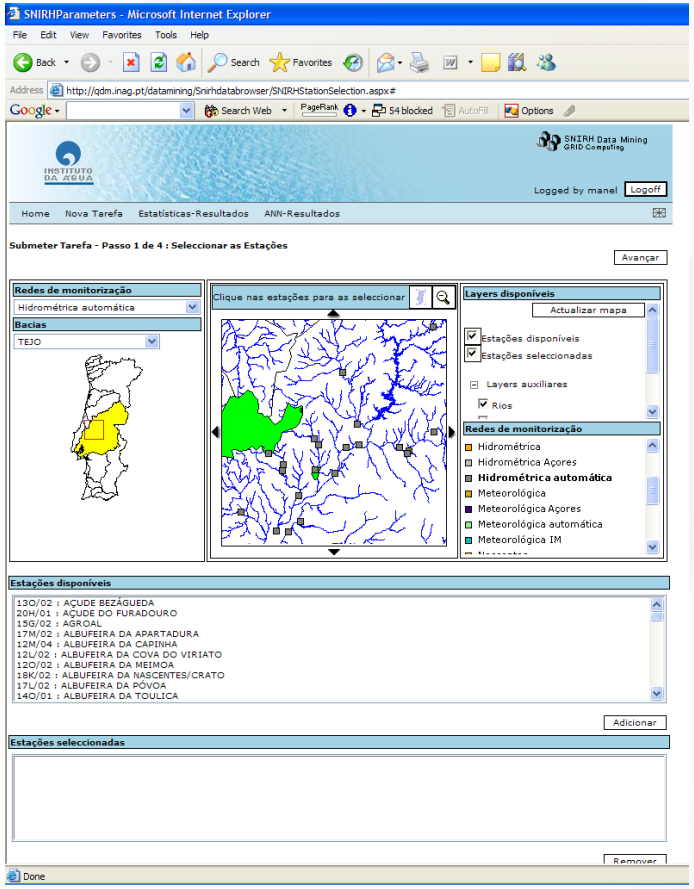
Modeling with insufficient knowledge about complex systems => 100% based on data: “you don’t have to know the underlying physics”.

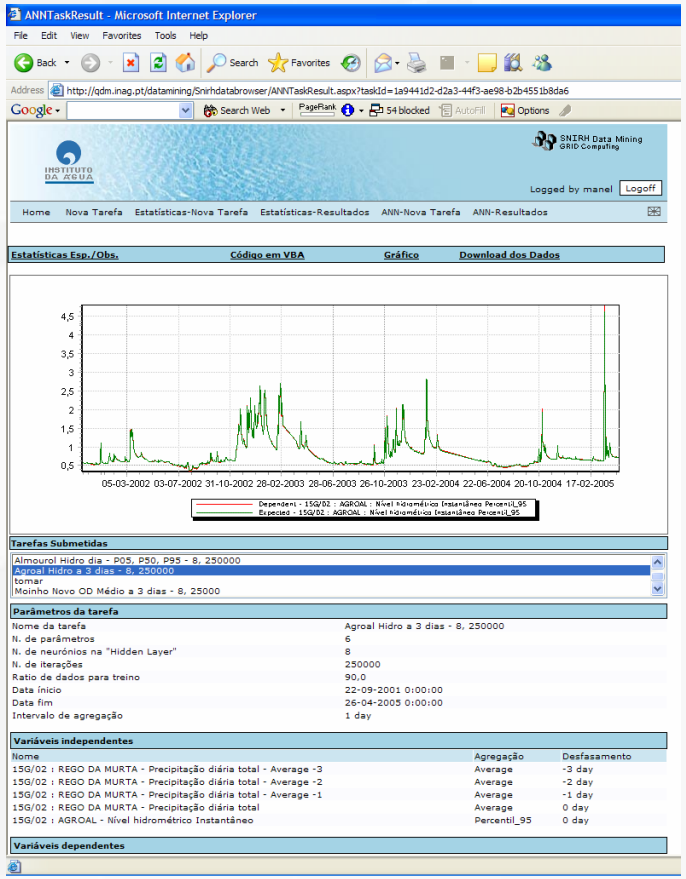
Data Driven Modeling Limitations

Modeling often requires computational muscle.

Some methods do not scale for large volumes of data.

SNIRH Data Mining





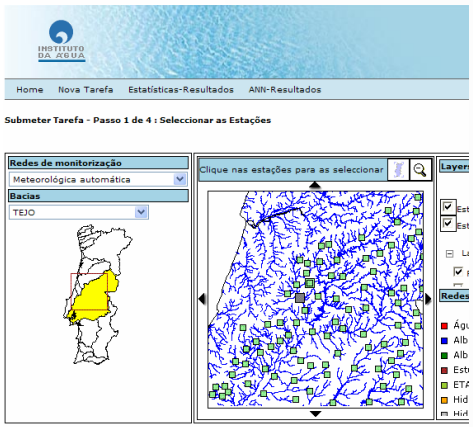
Online Data Mining Service (User's perspective)

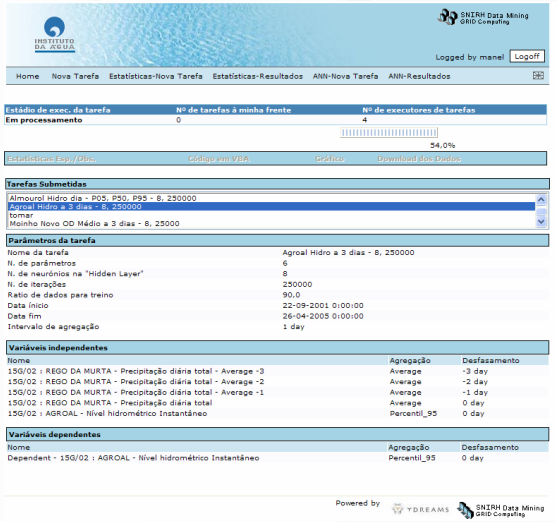
Meta-Information | Upload | Download | Web Services

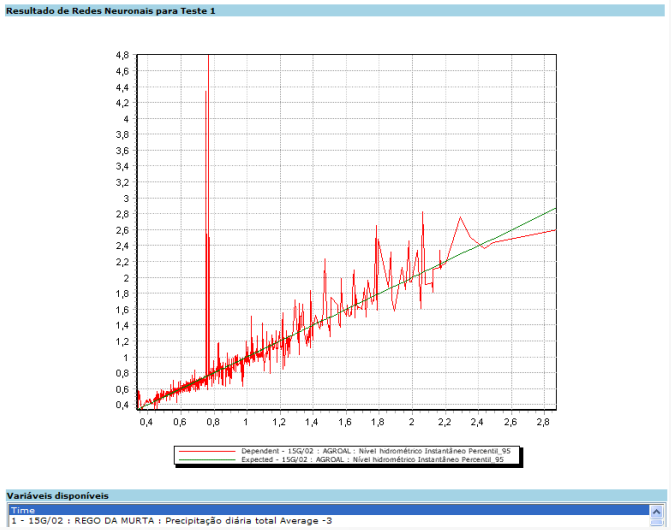
1 – Query existing data

2 – Task Upload

3 – Visualize results



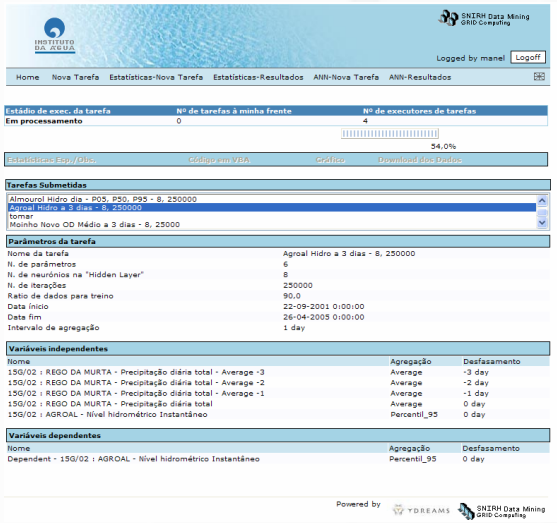




Online Data Mining Service (Server's perspective)

Meta-Information | Upload | Download | Web Services

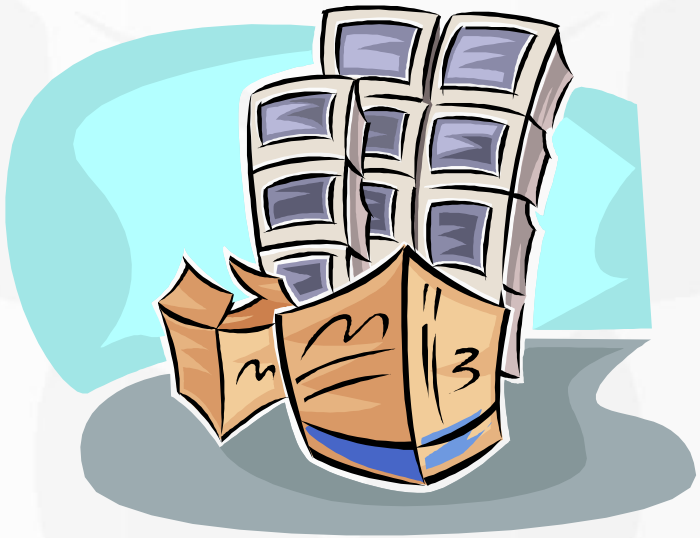
2 – Task Upload



Online Data Mining Service (Server's perspective)

Meta-Information	Upload	Download	Web Services

Online Data Mining Service (Server's perspective)



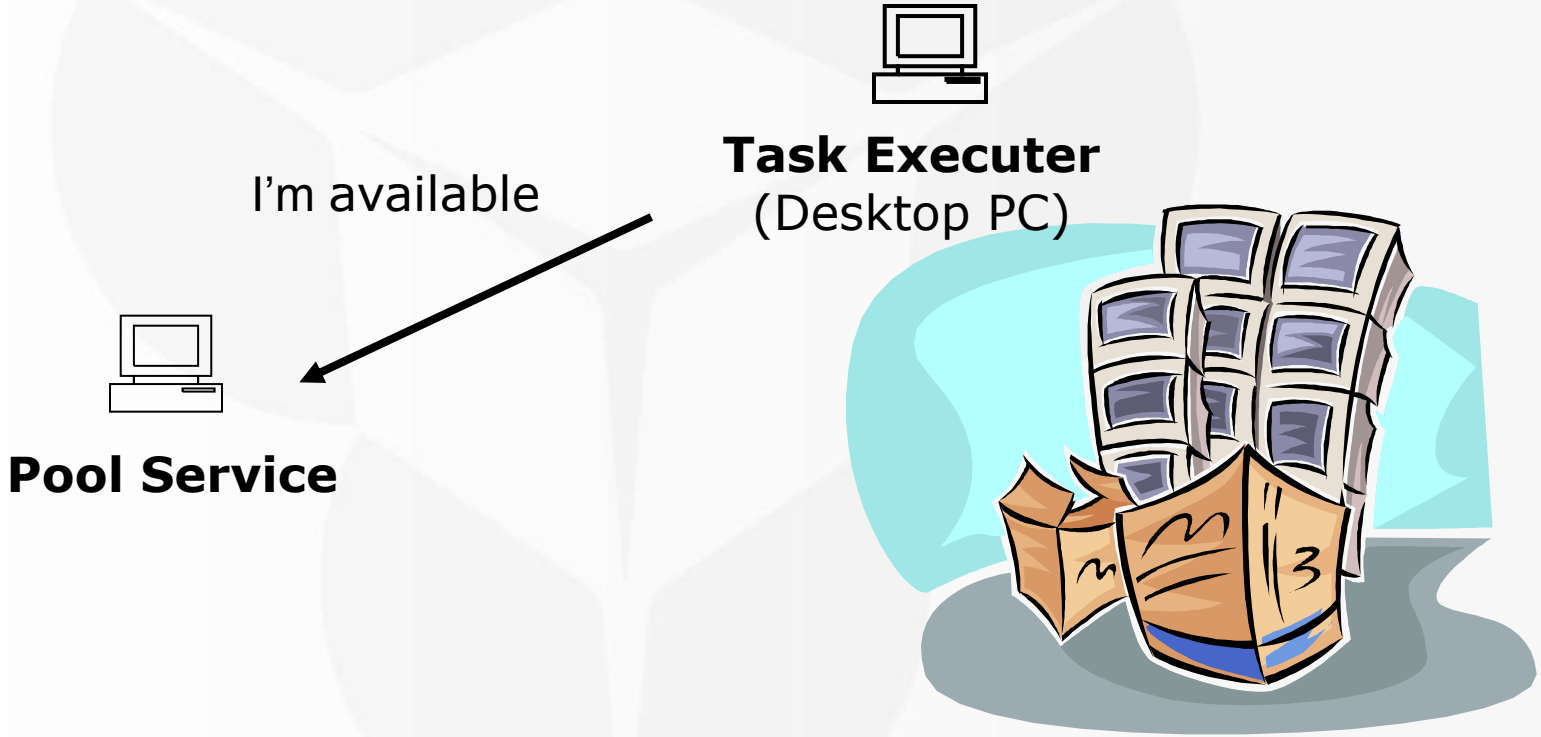
Task Result is queued

Task is queued

Executed

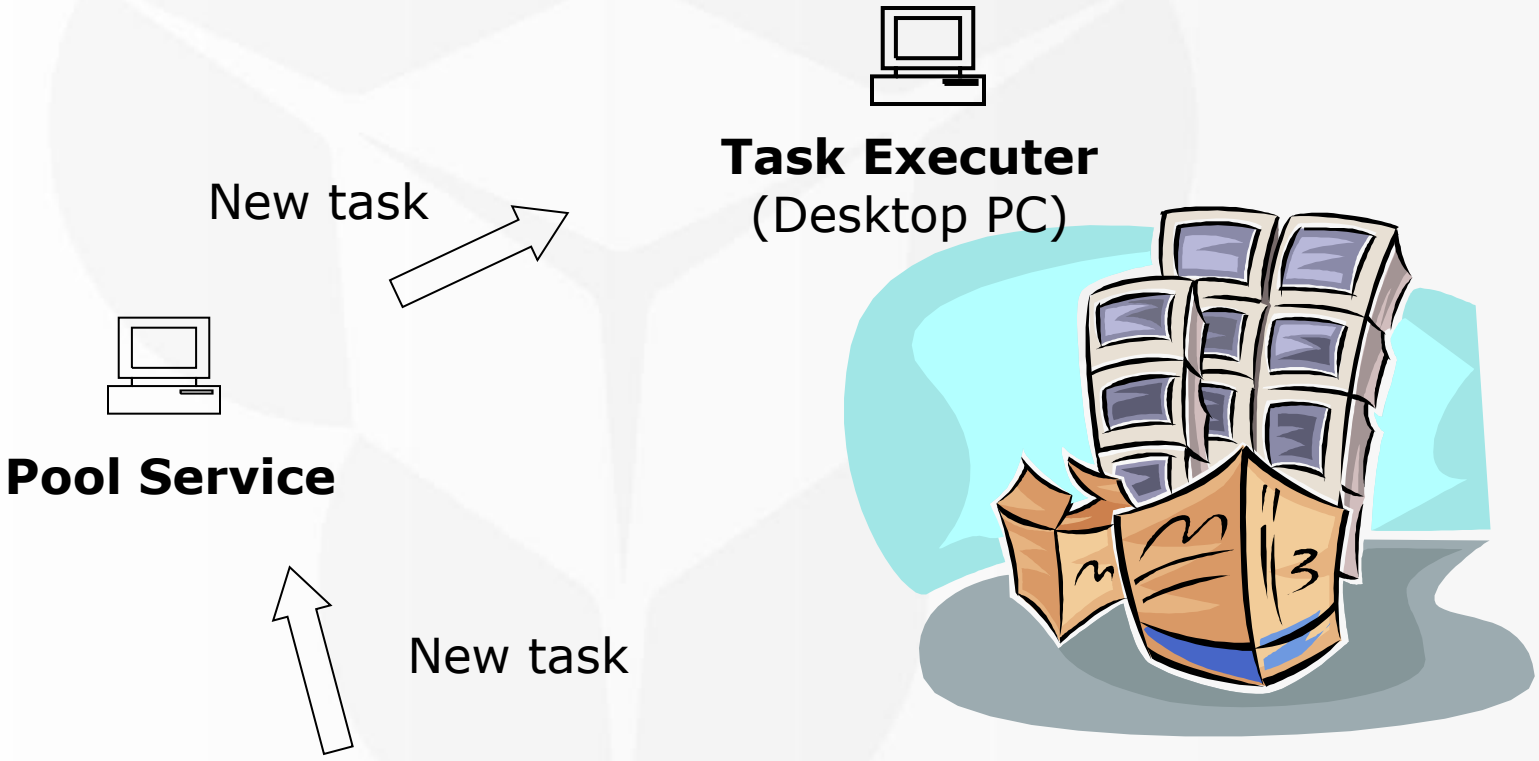
Meta-Information | Upload | Download | Web Services

Distributed task execution

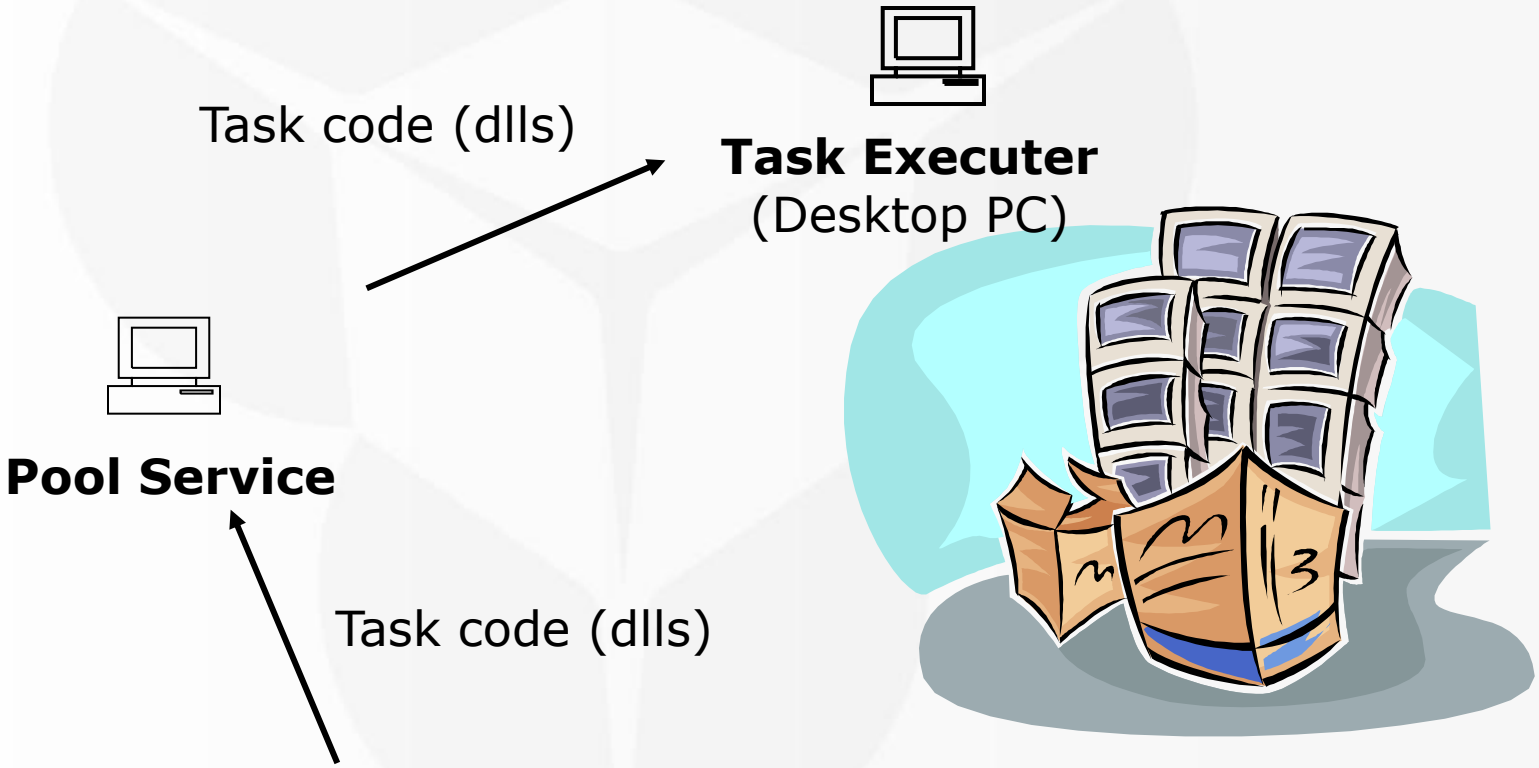


Meta-Information | Upload | Download | Web Services

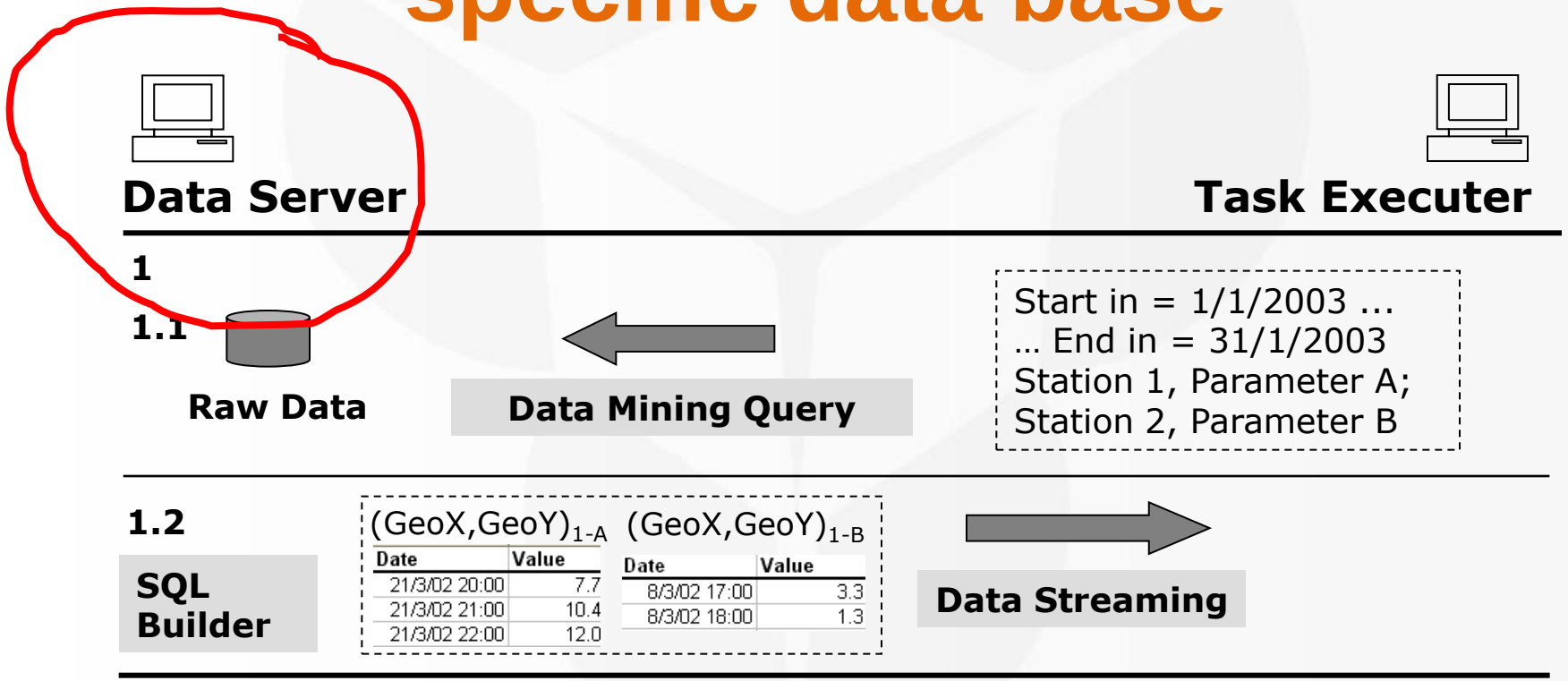
Distributed task execution



Distributed task execution



Customization : targeting a specific data base



Customization : targeting a specific data base

Data Server

Task Executer

2

Data Preparation

Date	Station 1, Parameter A	Station 2, Parameter B
	Percentile 95 per 24 Hours	Percentile 95 per 24 Hours
1/1/2003	1.1	4.3
2/1/2003	2.3	3.1

3

ANN Training

Station 2, Parameter B
Percentile 95 per 24 Hours
4.3
3.1

$=$ f(

Station 1, Parameter A
Percentile 95 per 24 Hours
1.1
2.3

)

Spatial Sound Data Mining, the project

Automatic Air Quality Index forecast 1 day ahead of 4 air pollutants (Nitrogen dioxide, ozone, sulphur dioxide, inhalable particles)

The prediction is made available to end users through mobile phones. By using **Location Based Services (LBS)**, the user receives the forecast for the closest monitoring station.

Spatial Sound Data Mining, the project

Applied to the **QualAr** database of the **Instituto do Ambiente**

Kick off of the Online Data Mining Services and a proof of concept

Spatial Sound Data Mining

QualAr Meta-Information Service

Auxiliary services

Sound  **Maps** 

Automatic Services (prediction of air quality index)

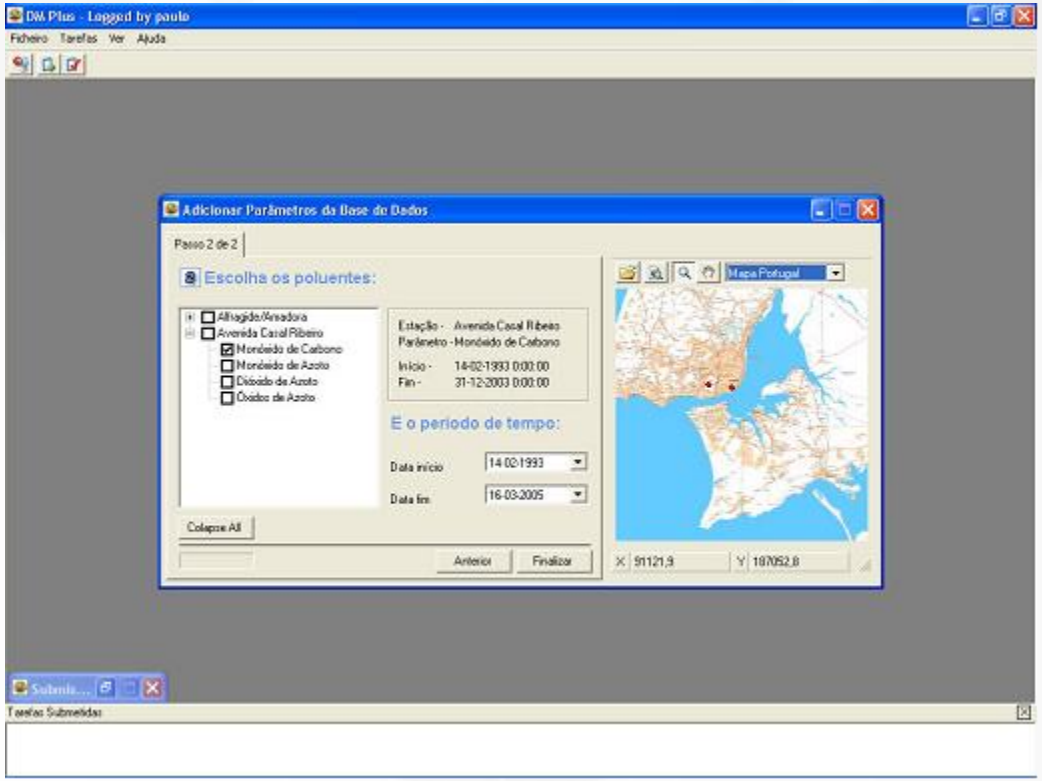
Data Mining Service

Client Applications

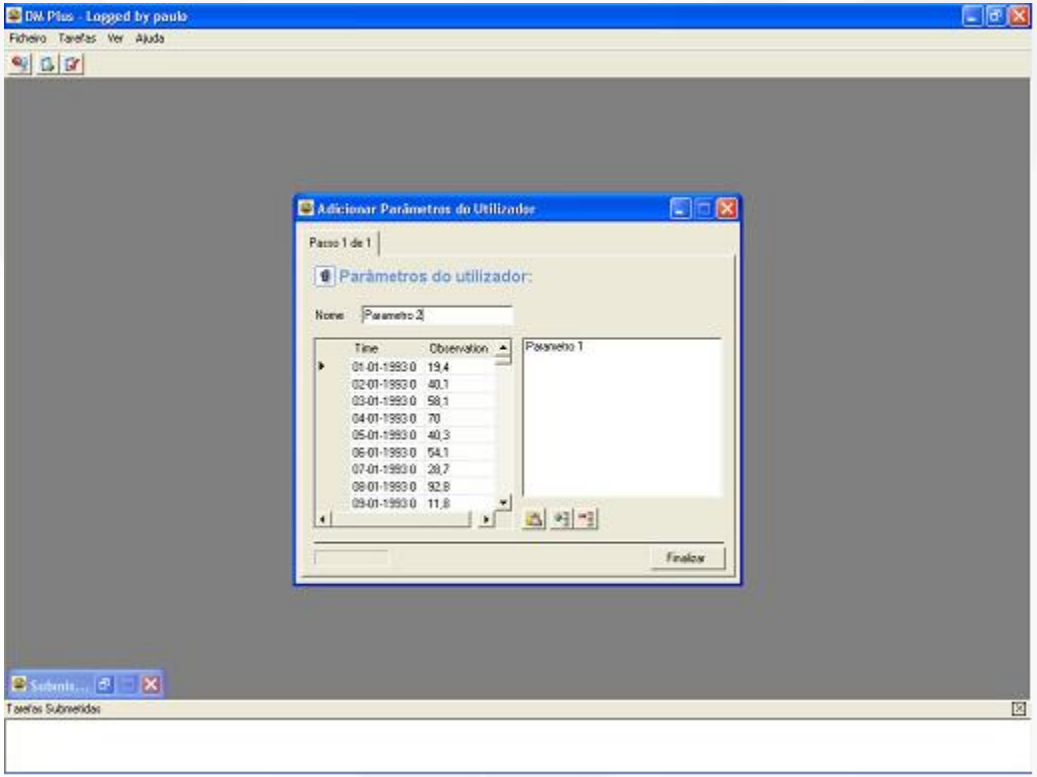
DM Plus QualAr

Automatic Forecast of the Air Quality Index

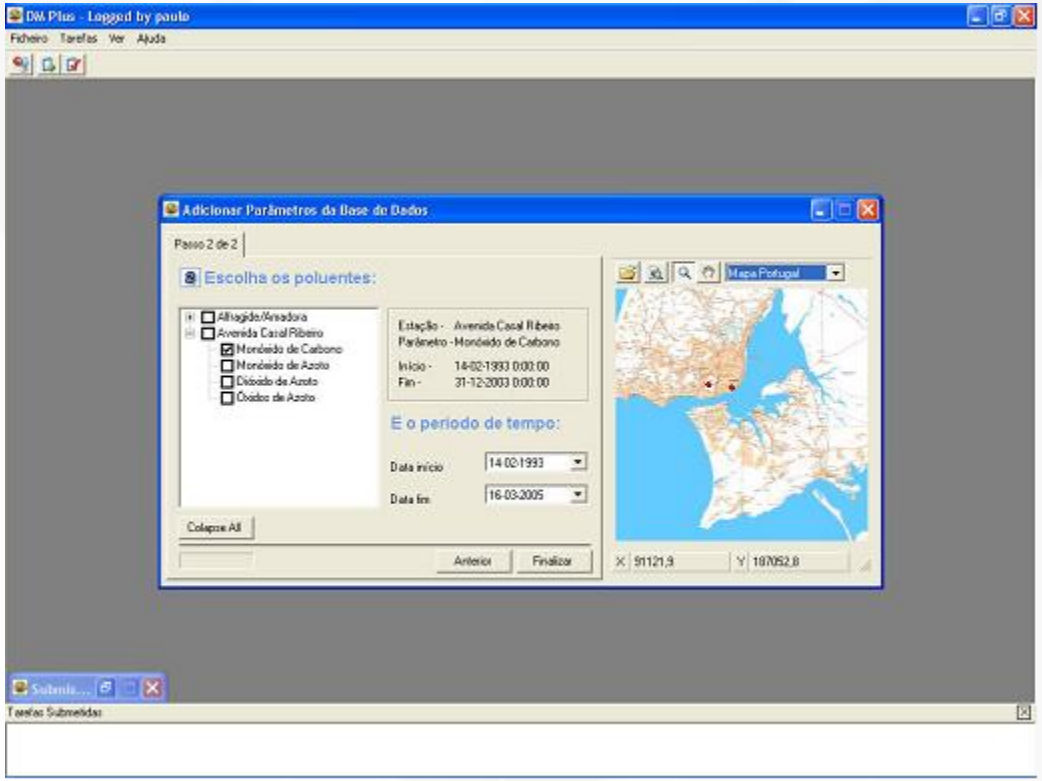
Spatial Sound Data Mining
DM Plus QualAr



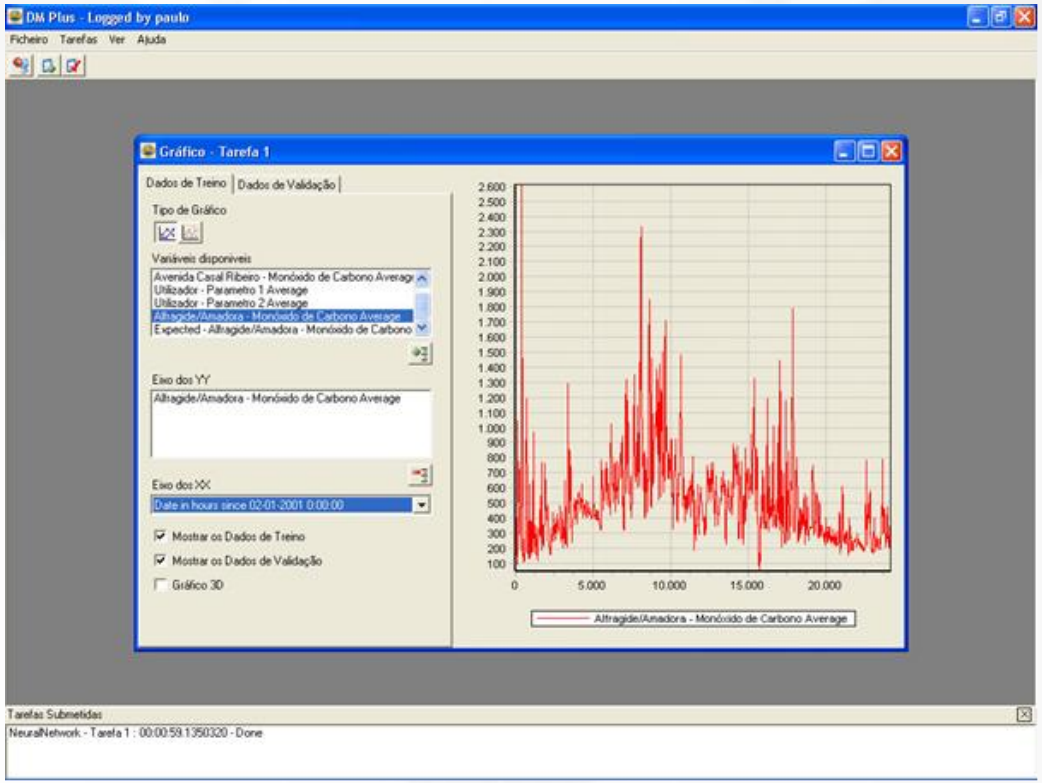
Spatial Sound Data Mining
DM Plus QualAr



Spatial Sound Data Mining
DM Plus QualAr



Spatial Sound Data Mining
DM Plus QualAr



Spatial Sound Data Mining

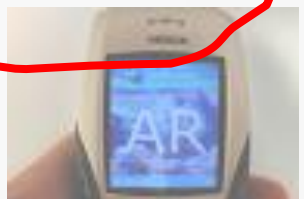
- After defining the most relevant Input parameters

Auxiliary services

Sound  **Maps** 

Automatic Services (prediction of air quality index)

Data Mining Service



Forecast Service uses automatic trial and error procedures to retrain the networks with new measured data.

Spatial Sound Data Mining Index representation

Limited graphical capabilities and processing constraints of the targeted platform (wireless mobile phone)

Sonification: use of non-speech audio to convey information

Spatial Sound Data Mining Index representation

Audio used to increase perception of the information:

- the air quality index and each pollutant are associated to a sound.

- *a melody is composed according to a composition rule that specifies the order and duration of each sound.*

Design challenge: create a sonification that conveys a relevant subset of the information presented in an acceptable time interval.

Spatial Sound Data Mining
Index representation

Server composed by two main units: Web Server Unit and Synthesis Unit

Client connects (java application) to the air quality service provider and requests information regarding the air quality status.

User listens to the sonification of both the index and its parameters

Marquis

European Air Quality database (6 Countries) and information services for the general public

The online data mining services will be used to interpolate empty records caused by anomalies in the monitoring stations

Conclusions

The online data mining services applied to environmental monitoring networks have shown to be a quite useful tool to associate with dynamic databases

The online data mining services, like in SNIRH Data Mining, is a new concept of data exploration

The online services to spatial analysis will happen in a near future

Contacts

manuel.costa@ydreams.com

SNIRH Data Mining

marco@inag.pt

Spatial Sound Data Mining Project

lmt@porto.ucp.pt, afonseca@igeo.pt