



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



www.ydreams.com

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 \Rightarrow 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

SNIRHParameters - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://qdm.inag.pt/datamining/Snrhdatabrowser/SNIRHStationSelection.aspx#

Google Search Web PageRank 54 blocked AutoFill Options

INSTITUTO DA ÁGUA SNIRH Data Mining GRID Computing

Logged by manel Logoff

Home Nova Tarefa Estatísticas-Resultados ANH-Resultados

Submeter Tarefa - Passo 1 de 4: Seleccionar as Estações

Avançar

Redes de monitorização

Hidrométrica automática

Bacias

TEJO

Clique nas estações para as seleccionar

Layers disponíveis

Actualizar mapa

☒ Estações disponíveis

☒ Estações seleccionadas

☐ Layers auxiliares

☒ Rios

Redes de monitorização

☐ Hidrométrica

☐ Hidrométrica Açores

☒ Hidrométrica automática

☐ Meteorológica

☐ Meteorológica Açores

☒ Meteorológica automática

☐ Meteorológica IM

Estações disponíveis

130/02 : AÇUDE BEZAGUEDA

20H/01 : AÇUDE DO FURADOURO

15G/02 : AGROAL

17M/02 : ALBUFEIRA DA APARTADURA

12M/04 : ALBUFEIRA DA CAPINHA

12L/02 : ALBUFEIRA DA COVA DO VIRIATO

12O/02 : ALBUFEIRA DA MEIMOA

18K/02 : ALBUFEIRA DA NASCENTES/CRATO

17L/02 : ALBUFEIRA DA PÓVOA

14O/01 : ALBUFEIRA DA TOULICA

Adicionar

Estações seleccionadas

Remover

Done

ANNTaskResult - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://qdm.inag.pt/datamining/Snrhdatabrowser/ANNTaskResult.aspx?taskId=1a9441d2-d2a3-44f3-ae98-62b4551b8da6

Google Search Web PageRank 54 blocked AutoFill Options

INSTITUTO DA ÁGUA SNIRH Data Mining GRID Computing

Logged by manel Logoff

Home Nova Tarefa Estatísticas-Nova Tarefa Estatísticas-Resultados ANH-Nova Tarefa ANH-Resultados

Estatísticas Esp./Obs. Código em VBA Gráfico Download dos Dados

Gráfico

Dependente - 15G/02 : AGROAL : Nivel Hidrométrico Instantâneo Percentil_95

Expediente - 15G/02 : AGROAL : Nivel Hidrométrico Instantâneo Percentil_95

Tarefas Submetidas

Almoural Hidro dia - P05, P50, P95 - 8, 250000

Proprietário: S. Edm. Sd - 250000

tomar

Moinho Novo OD Médio a 3 dias - 8, 25000

Parâmetros da tarefa

Nome da tarefa Agrol Hidro a 3 dias - 8, 250000

N. de parâmetros 6

N. de neurónios na "Hidden Layer" 8

N. de iterações 250000

Ratio de dados para treino 90.0

Data início 22-09-2001 0:00:00

Data fim 26-04-2009 0:00:00

Intervalo de agregação 1 day

Variáveis independentes

Nome	Agregação	Desfasamento
15G/02 : REGO DA MURTA - Precipitação diária total - Average -3	Average	-3 day
15G/02 : REGO DA MURTA - Precipitação diária total - Average -2	Average	-2 day
15G/02 : REGO DA MURTA - Precipitação diária total - Average -1	Average	-1 day
15G/02 : REGO DA MURTA - Precipitação diária total	Average	0 day
15G/02 : AGROAL - Nivel hidrométrico Instantâneo	Percentil_95	0 day

Variáveis dependentes



www.ydreams.com

Online Data Mining Service (User's perspective)

Meta-Information

Upload

Download

Web Services

1 – Query
existing data

2 – Task Upload

3 – Visualize results

INSTITUTO DA ÁGUA

Home Nova Tarefa Estatísticas-Resultados ANI-Resultados

Submeter Tarefa - Passo 1 de 4 : Selecionar as Estações

Redes de monitorização
Meteorológica automática

Bacias
TEJO

Clique nas estações para as seleccionar

Layer

Redes

- Ag
- Alb
- Est
- Li
- mi
- Hid

INSTITUTO DA ÁGUA

SNRM Data Mining
SRD Consulting

Logged by manel Logout

Home Nova Tarefa Estatísticas-Nova Tarefa Estatísticas-Resultados ANI-Nova Tarefa ANI-Resultados

Estado da exec. da tarefa 0 Nº de tarefas à minha frente 1 Nº de executores de tarefas 1

Em processamento 0 34.0%

Estatísticas Esp/Disc. Código em YMA Gráfico Download dos dados

Tarefas Submetidas

Almoural Hidro dia - PDS, P50, P95 - 8, 250000
Agroal Hidro a 3 dias - 8, 250000
Moimho Novo OD Médio a 3 dias - 8, 25000

Parâmetros da tarefa

Nome da tarefa	Agroal Hidro a 3 dias - 8, 250000
N.º de parâmetros	6
N.º de neurónios na "hidden layer"	8
N.º de iterações	250000
Ratão de dados para treino	90.0
Data início	22-09-2001 0:00:00
Data fim	26-04-2005 0:00:00
Intervalo de agregação	1 day

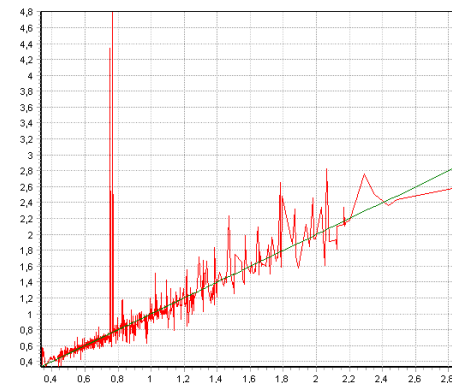
Variáveis independentes

Nome	Agregação	Desfasamento
150/02 : REGO DA MURTA - Precipitação diária total - Average -3	Average	-3 day
150/02 : REGO DA MURTA - Precipitação diária total - Average -2	Average	-2 day
150/02 : REGO DA MURTA - Precipitação diária total - Average -1	Average	-1 day
150/02 : REGO DA MURTA - Precipitação diária total - Average	Average	0 day
150/02 : AGROAL - Nível hidrométrico Instantâneo	Percentil_95	0 day

Variáveis dependentes

Nome	Agregação	Desfasamento
Dependent - 150/02 : AGROAL - Nível hidrométrico Instantâneo	Percentil_95	0 day

Resultado de Redes Neurais para Teste 1



Variáveis disponíveis

Time
1 - 150/02 : REGO DA MURTA : Precipitação diária total Average -3

Powered by YDREAMS SNRM Data Mining SRD Consulting



www.ydreams.com

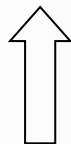
Online Data Mining Service (Server's perspective)

Meta-Information

Upload

Download

Web Services



2 – Task Upload

INTELM Data Mining GRID Computing

Home Nova Tarefa Estatísticas-Nova Tarefa Estatísticas-Resultados ANM-Nova Tarefa ANM-Resultados

Logged by manel Logout

Estado da exec. da tarefa

Em processamento	Nº de tarefas à minha frente	Nº de executores de tarefas
0	4	

Progresso 34.0%

Tarefas Submetidas

Tarefa	Estado
Almourol Hidro dia - P05, P50, P95 - 8, 250000	Em processamento
Almourol Hidro dia - P05, P50, P95 - 8, 250000	Em processamento
Tomar	Em processamento
Moinho Novo OD Médio a 3 dias - 8, 25000	Em processamento

Parâmetros da tarefa

Nome da tarefa	Agenda Hidro a 3 dias - 8, 250000
N. de parâmetros	6
N. de neurónios na "Hidden Layer"	8
N. de iterações	250000
Ratio de dados para treino	90.0
Data início	22-09-2001 0:00:00
Data fim	26-04-2003 0:00:00
Intervalo de agregação	1 day

Variáveis independentes

Nome	Agregação	Desfasamento
150/02 : REGO DA MURTA - Precipitação diária total - Average -3	Average	-3 day
150/02 : REGO DA MURTA - Precipitação diária total - Average -2	Average	-2 day
150/02 : REGO DA MURTA - Precipitação diária total - Average -1	Average	-1 day
150/02 : REGO DA MURTA - Precipitação diária total	Average	0 day
150/02 : AGROAL - Nivel hidrométrico Instantâneo	Percentil_95	0 day

Variáveis dependentes

Nome	Agregação	Desfasamento
Dependent - 150/02 : AGROAL - Nivel hidrométrico Instantâneo	Percentil_95	0 day

Powered by YDREAMS INTELM Data Mining GRID Computing



www.ydreams.com

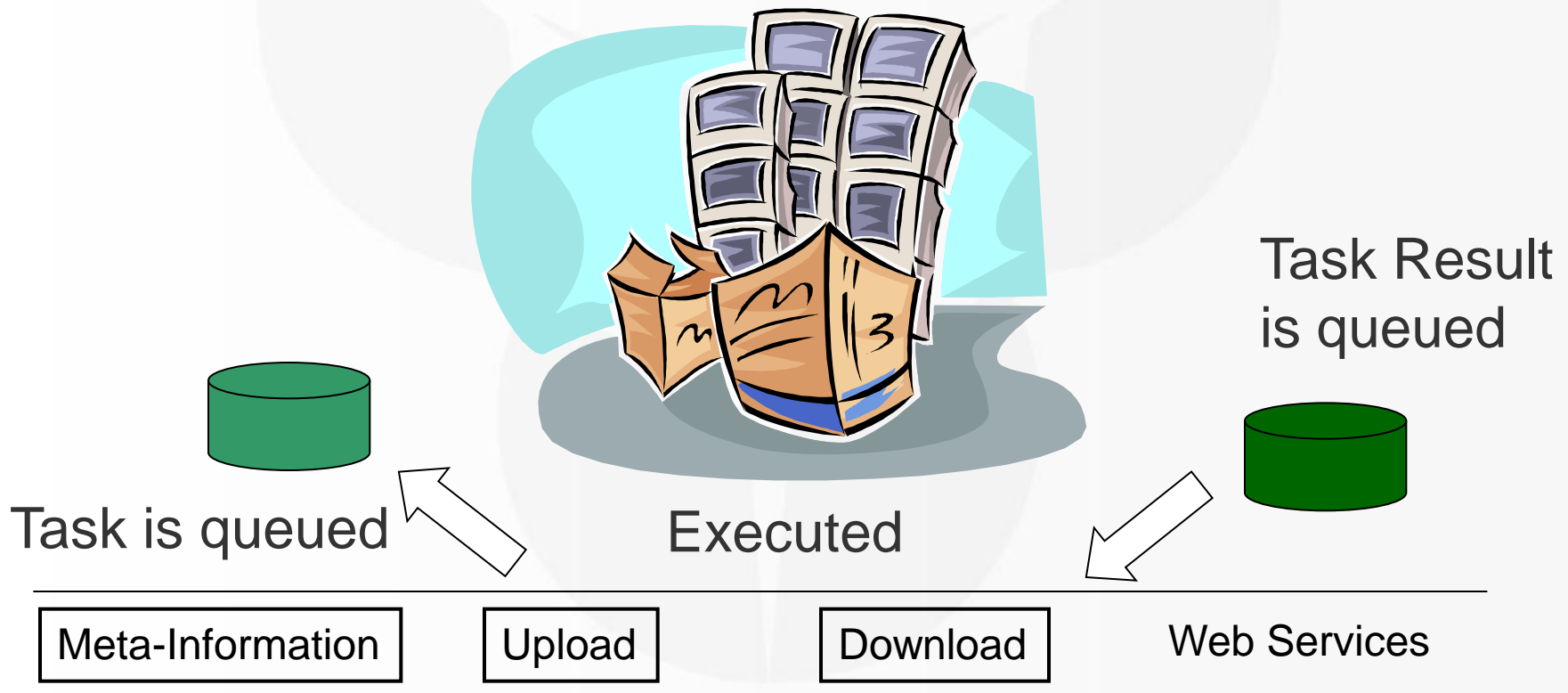
Online Data Mining Service (Server's perspective)

Meta-Information	Upload	Download	Web Services
------------------	--------	----------	--------------



www.ydreams.com

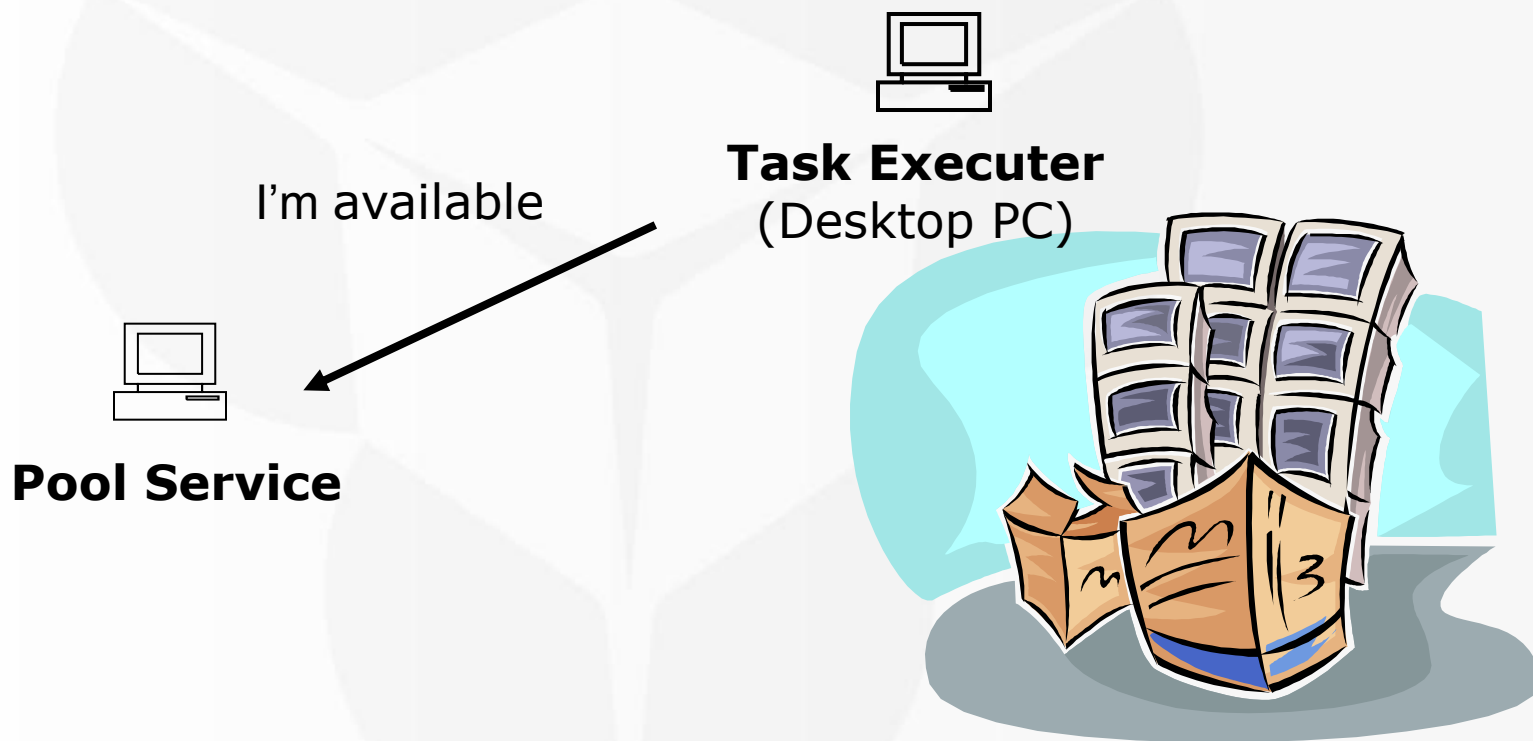
Online Data Mining Service (Server's perspective)





www.ydreams.com

Distributed task execution



Meta-Information

Upload

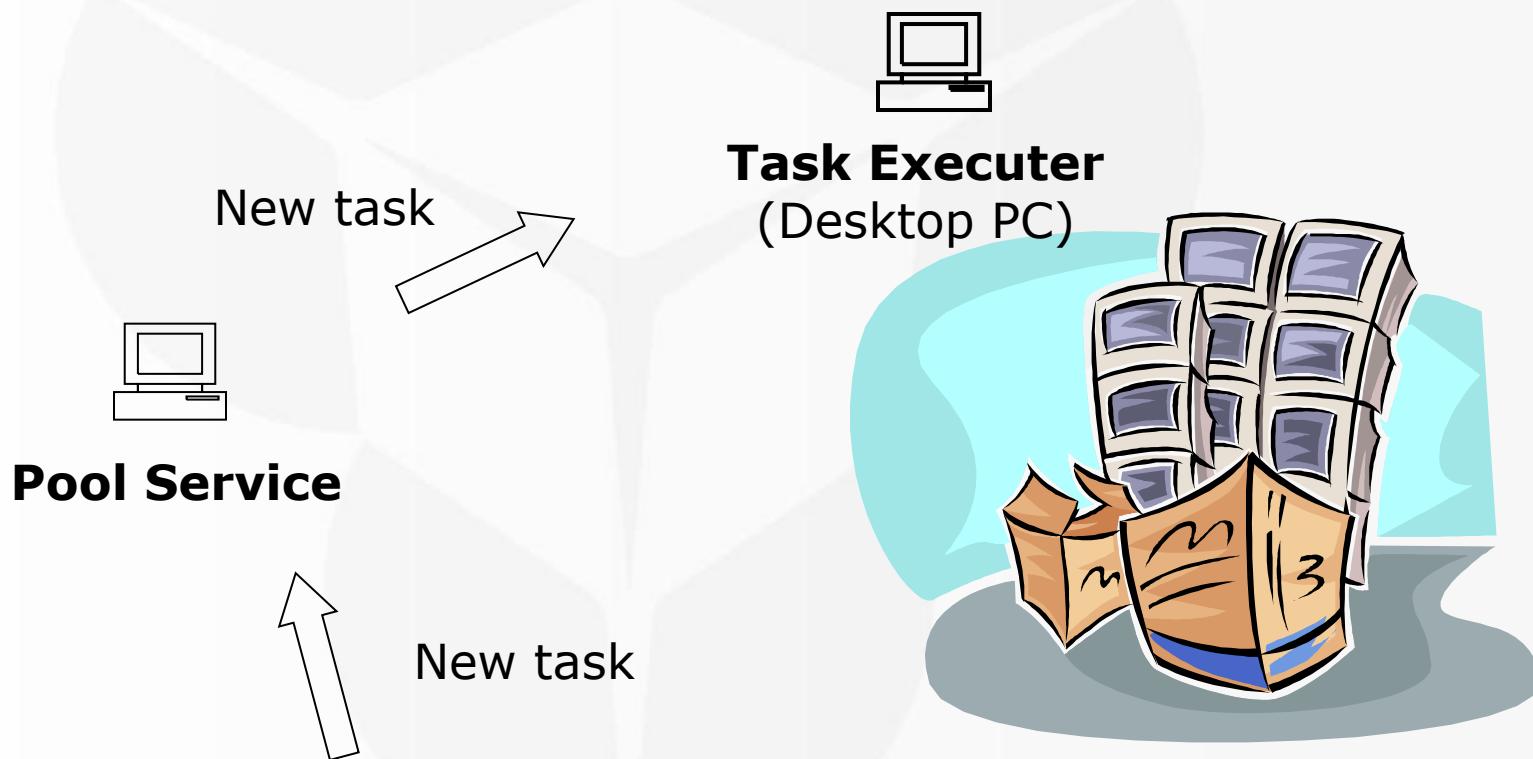
Download

Web Services



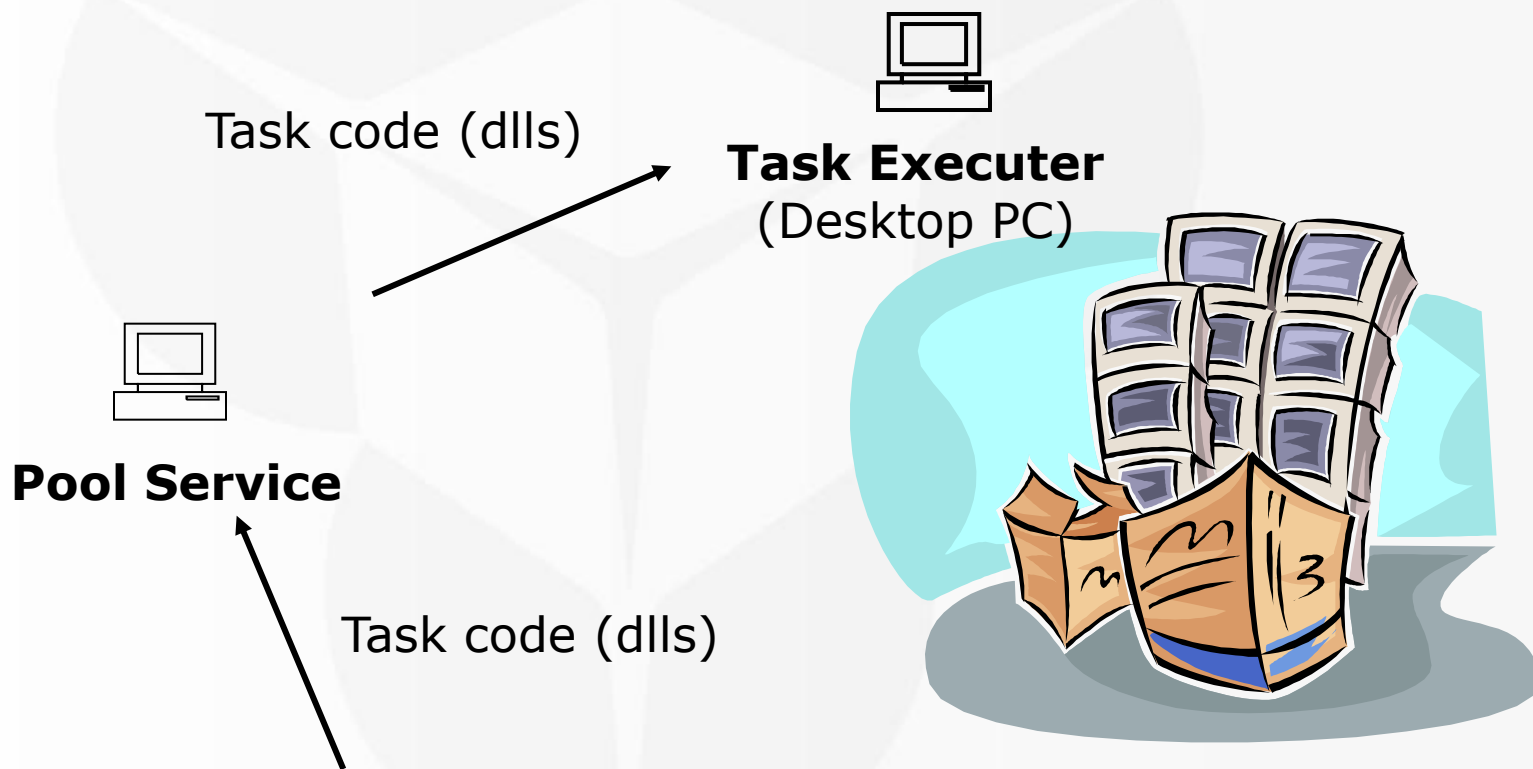
www.ydreams.com

Distributed task execution



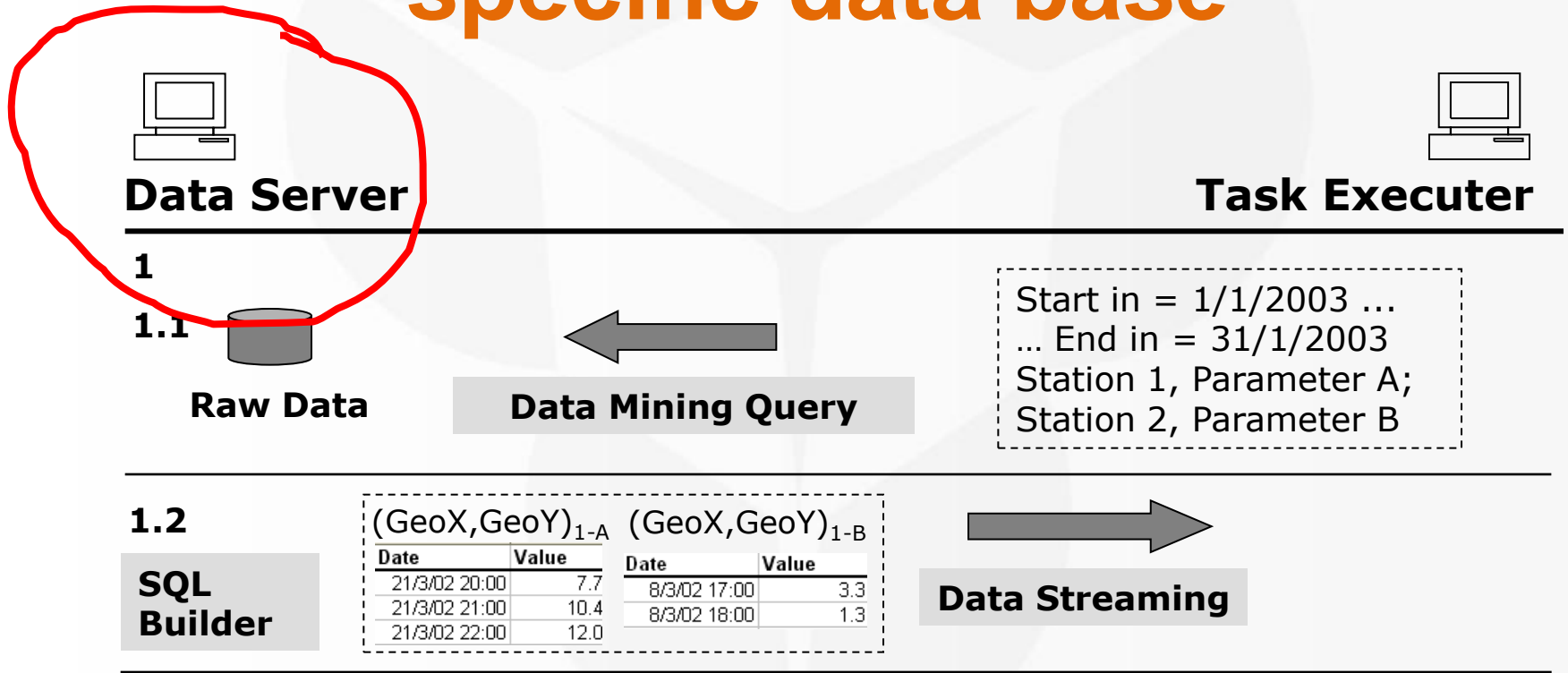


Distributed task execution



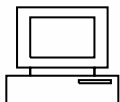


Customization : targeting a specific data base





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

$$\begin{array}{|c|c|} \hline \text{Station 2, Parameter B} \\ \hline \text{Percentile 95 per 24 Hours} \\ \hline 4.3 \\ \hline 3.1 \\ \hline \end{array} = f\left(\begin{array}{|c|c|} \hline \text{Station 1, Parameter A} \\ \hline \text{Percentile 95 per 24 Hours} \\ \hline 1.1 \\ \hline 2.3 \\ \hline \end{array}\right)$$



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.



www.ydreams.com

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



www.ydreams.com

Spatial Sound Data Mining

QualAr Meta-Information Service

Auxiliary services

Sound



Maps



Data Mining Service

**Automatic Services
(prediction of air quality index)**

Client Applications



DM Plus QualAr



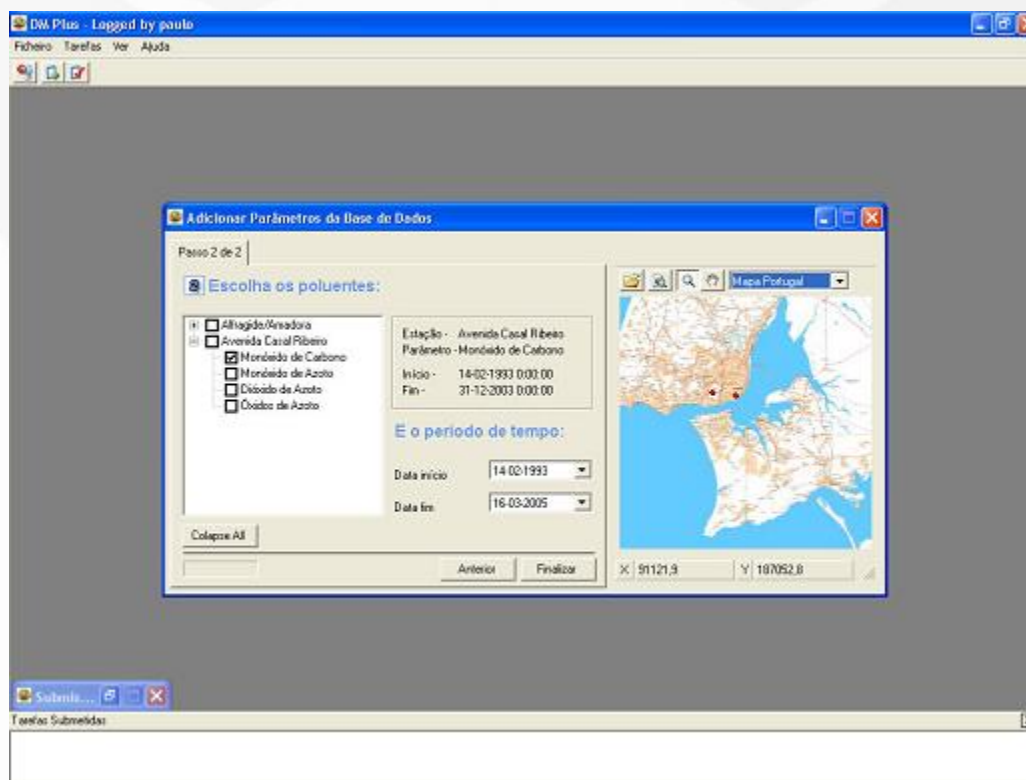
**Automatic Forecast of
the Air Quality Index**



www.ydreams.com

Spatial Sound Data Mining

DM Plus QualAr

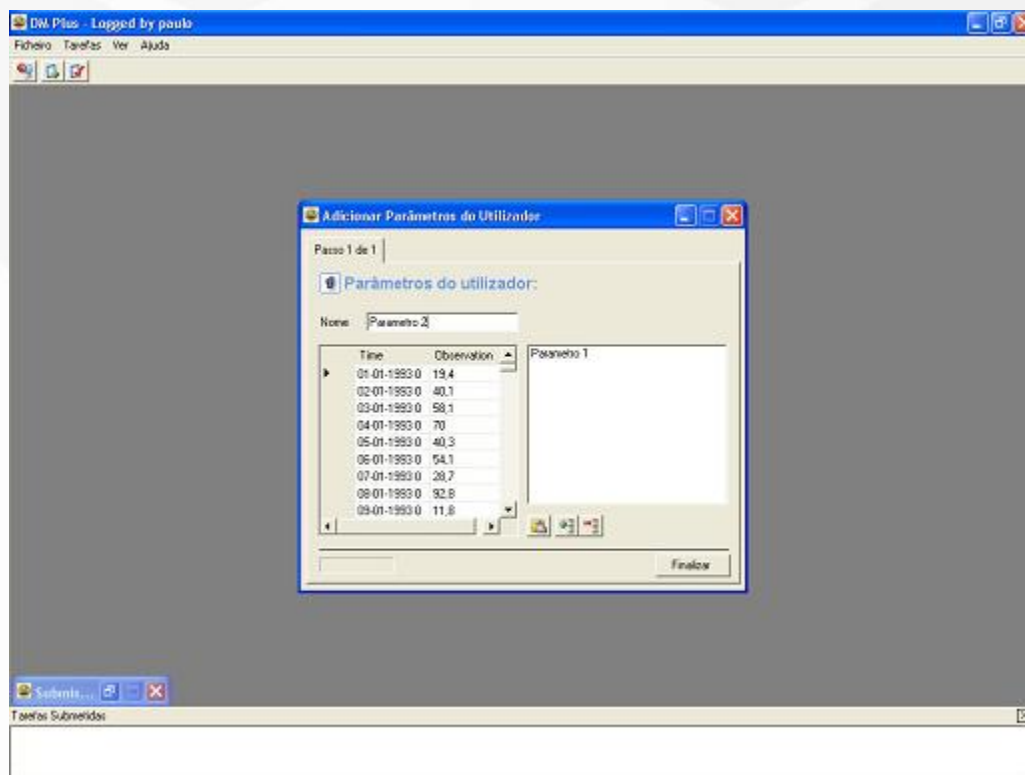




www.ydreams.com

Spatial Sound Data Mining

DM Plus QualAr

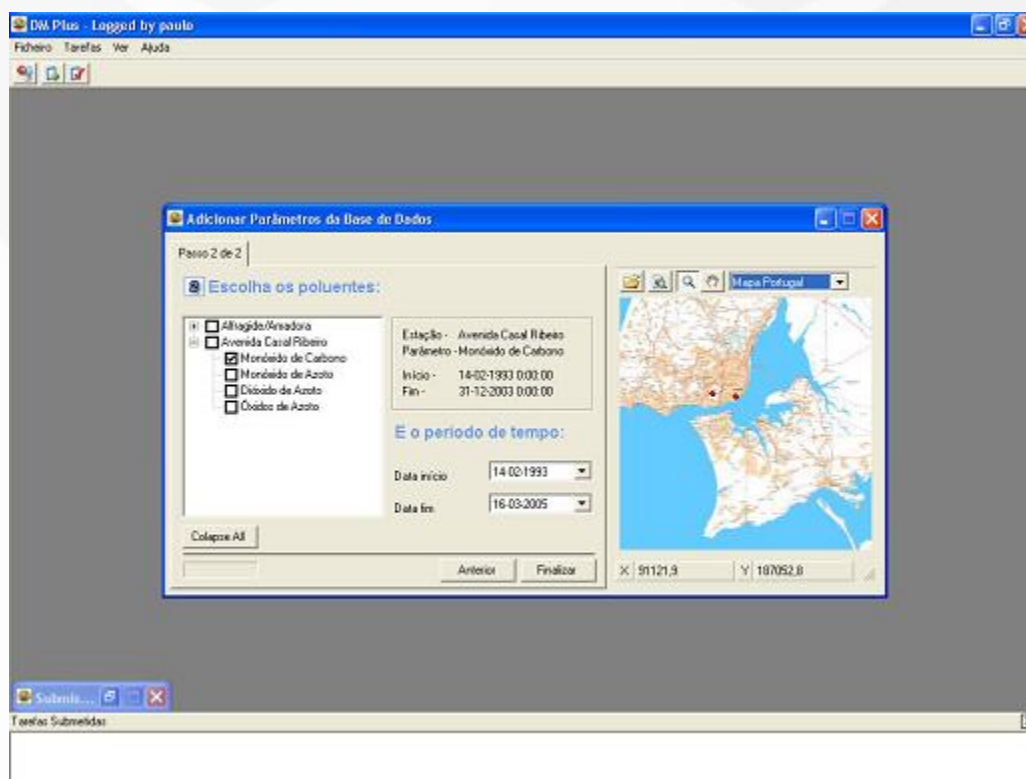




www.ydreams.com

Spatial Sound Data Mining

DM Plus QualAr

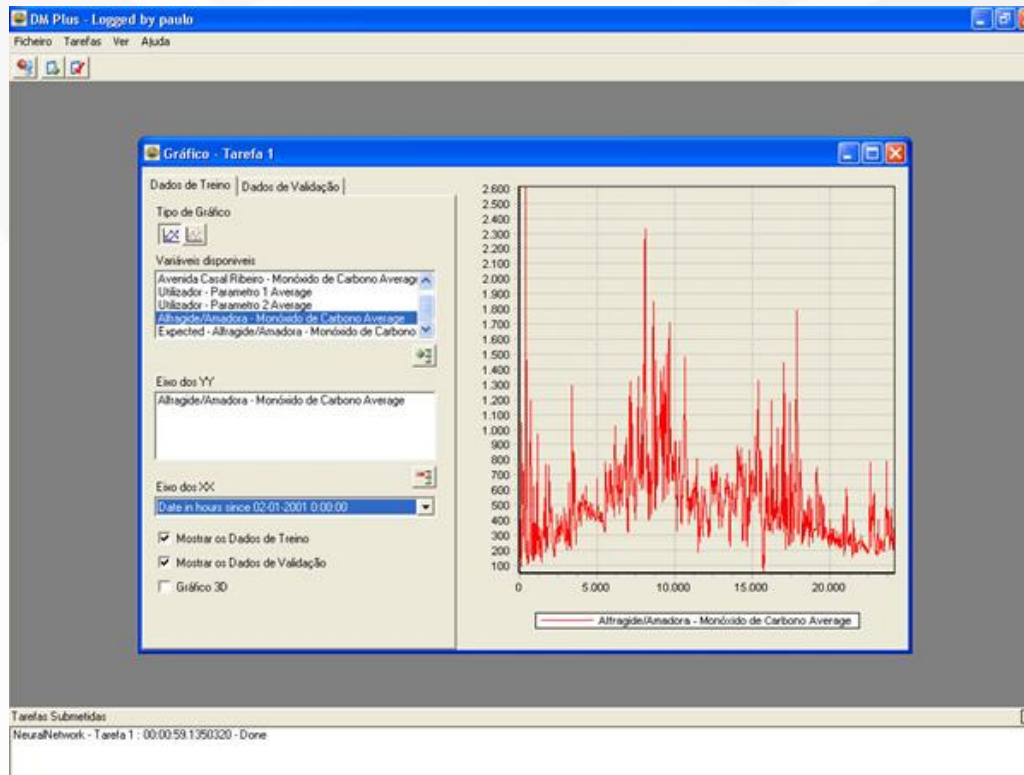




www.ydreams.com

Spatial Sound Data Mining

DM Plus QualAr





www.ydreams.com

Spatial Sound Data Mining

After defining the most relevant
Input parameters

**Auxiliary
services**

Sound



Maps



**Data Mining
Service**

**Automatic Services
(prediction of air quality
index)**



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.



www.ydreams.com

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



www.ydreams.com

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



www.ydreams.com

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



www.ydreams.com

Contacts

manuel.costa@ydreams.com

SNIRH Data Mining

marco@inag.pt

Spatial Sound Data Mining Project

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