Addressing High Resolution Modelling over different computing infrastructures (HPC, HTC, Cloud)

Fernando Aguilar
aguilarf@ifca.unican.es
Instituto de Física de Cantabria (IFCA)
Santander - Spain

Content

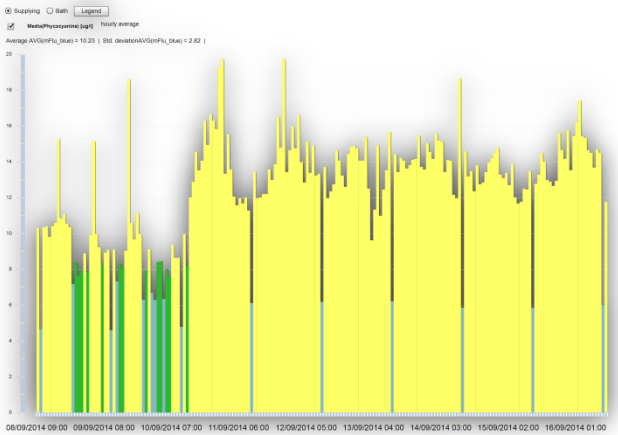
1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

Content

1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

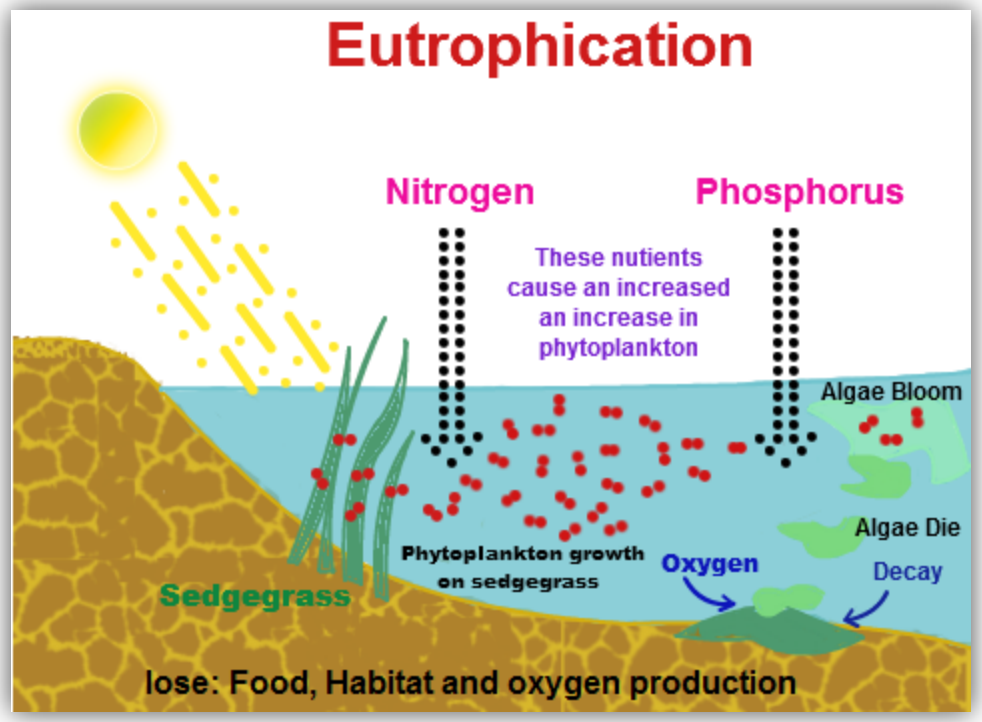
Introduction

- Framework: Collaboration within European LIFE+ project (ROEM+). SME Ecohydros.
- Reservoir hydrodinamic and Water Quality modelling. Cuerda del Pozo: water supply, water activities.
- Previous work
  - Platform takes data from water: physical, chemical, biological, etc. Allows to know water status (data taken since 2010 aprox.)
  - Data visualization tool. Aims to alert authorities when the water quality is under the limits.



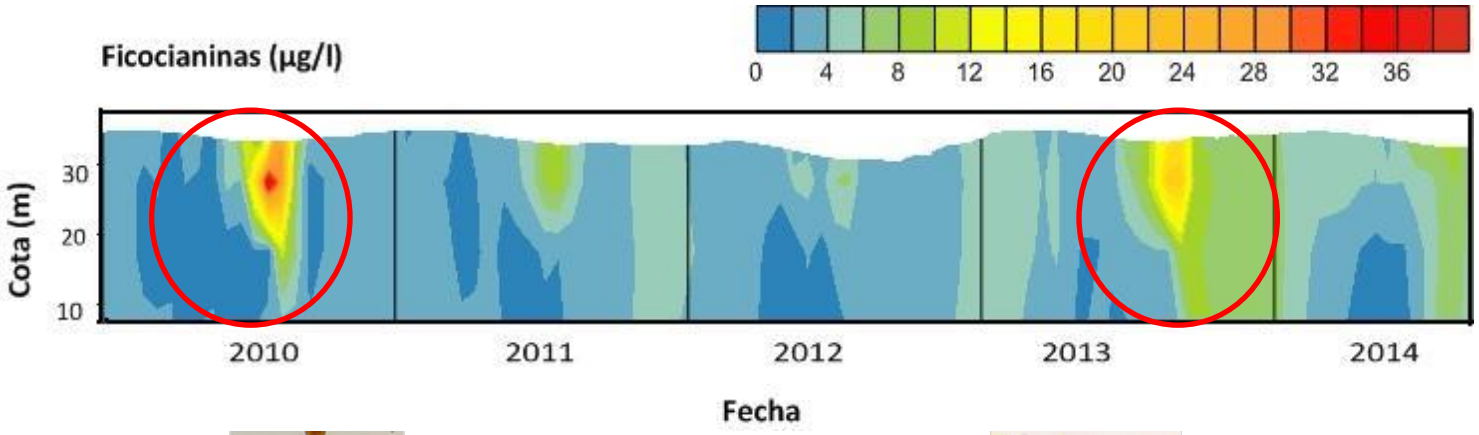
Introduction

- One more step: knowing before an event happens the status of the water using modelling tools (Delft3D used in cloud).
- Goal: alert authorities not only in real time but before.
- The main problem is eutrophication:



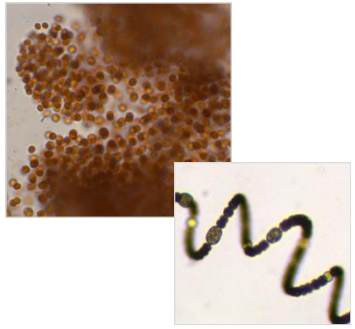
Introduction

Within 5 years of continuous monitoring in CdP Reservoir, this is the cyanobacterias concentration close to the dam:



2010 - 2011
Dolichospermum planctonicum
Aphanizomenon flos-aquae

2013
Dolichospermum crassum
Colonias Woronichinia naegeliana



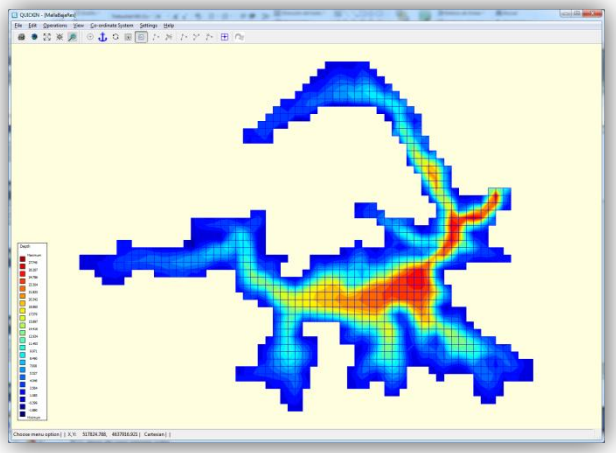
2014
Colonias Microcystis novacekii
Dolichospermum crassum

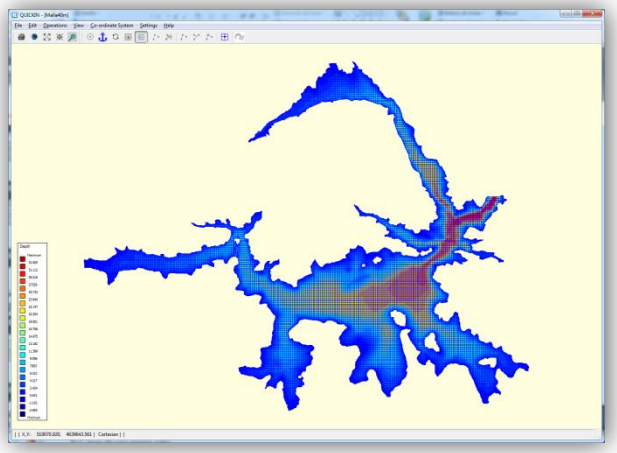
Content

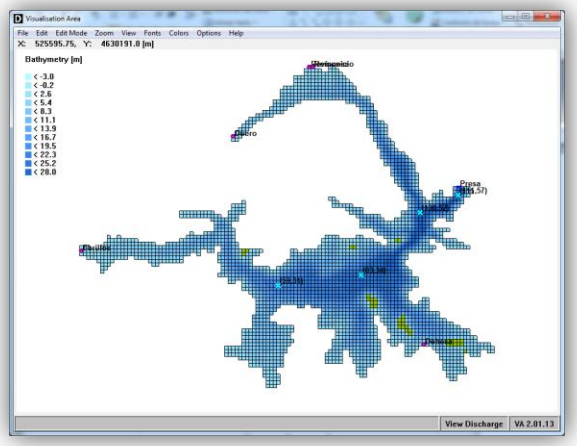
1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

Hydrodynamic Modelling

- Delft3D-FLOW: Different resolutions (Bathymetry, 5-40m horizontally, 0.5-3m vertically, 35 layers). Z-model.
- Number of input parameters:
  - Tributaries/Initial conditions: flow, temperature, salinity.
  - Meteo: Rain, air temp, humidity, solar radiation, wind.

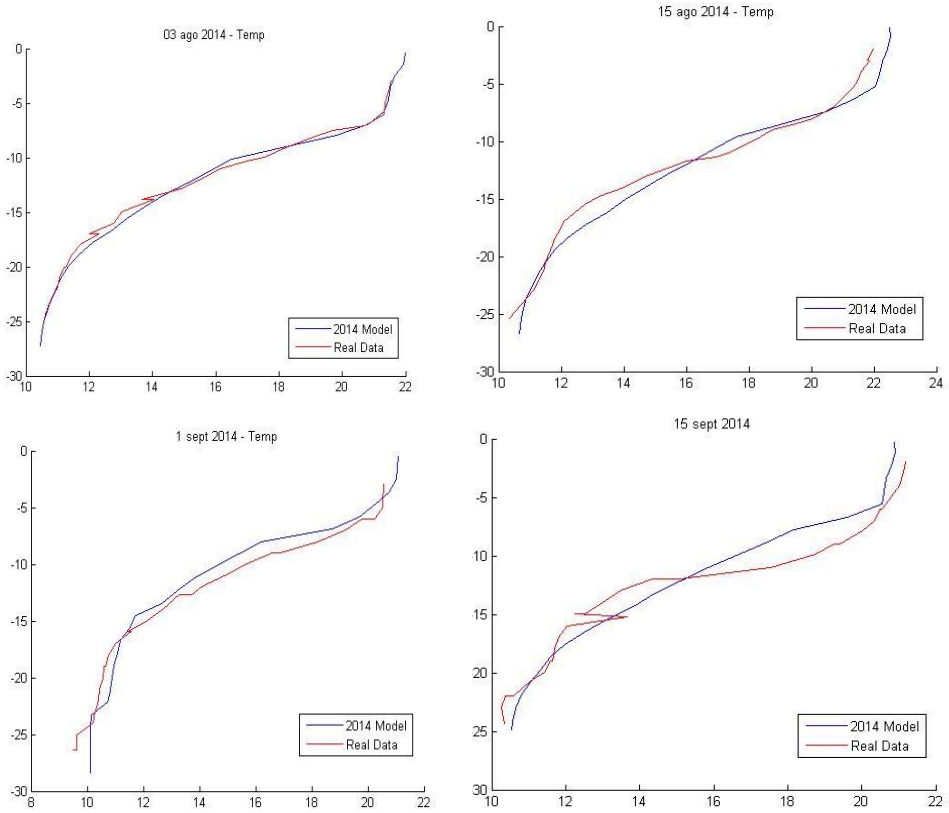


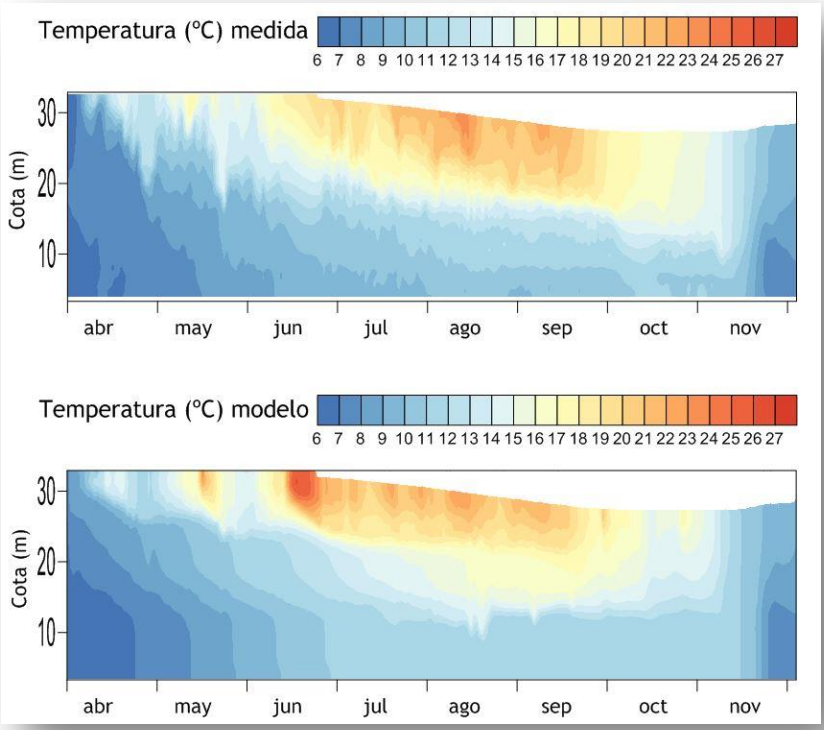




Hydrodynamic Modelling

- Goal: Reproduce thermocline and water level.
- Not easy, but very good results.



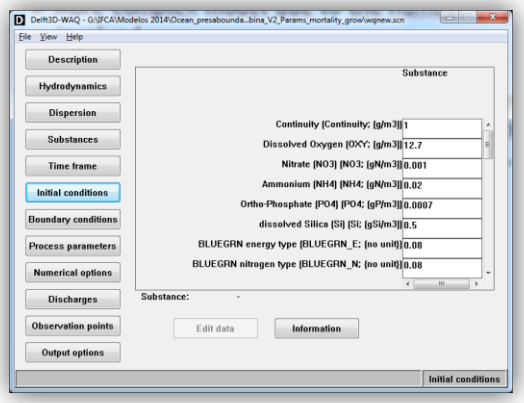


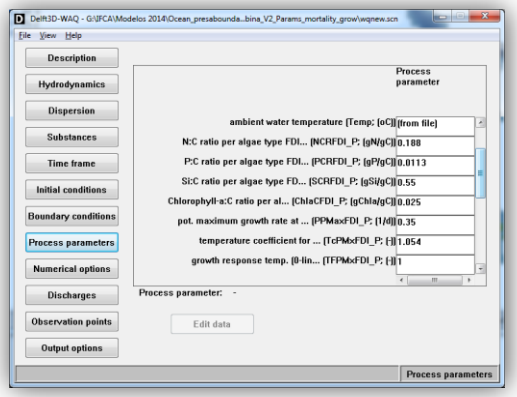
Content

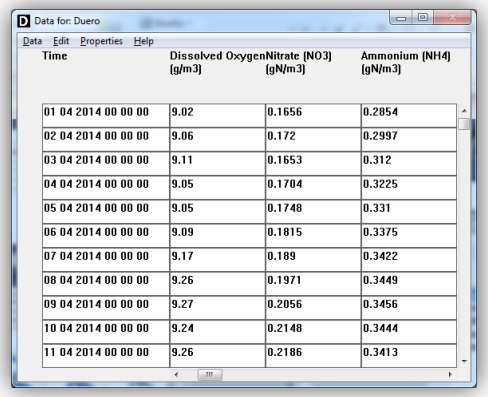
1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

Water Quality Modelling

- More complex model due to the number of processes involved.
- Goal: reproduce algae bloom.
- Input: hydrodynamic output, nutrient concentrations (initial/tributaries), initial algae concentrations, sediments, other coefficients/ratios (mortality, growth, ratio chl/C, etc.).







Water Quality Modelling

- We tried to model it increasingly, adding parameters one by one: Continuity>Oxygen>Nutrients…
- The model is not validated yet. Problems found.
  - Oxigen is not well distributed in the water column.
  - Nutrients (Phosphatus, Nitrates) decreasing (and dissapear).
  - Algae trend to die.
- Problems with continuity solved (time step adjustment).
- Tests in forced scenarios:
  - +50% nutrients
  - +90% nutrients
  - Algae phenotypes checked.
- Any ideas? Any Delf3D Algae bloom expert?

Content

1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

Infrastructure

- Low resolution works in standard PCs. For high resolution more powerful resources are needed (100x100 m, 30-35 vertical layers).
- “Z model” is not parallelizable itself. Parallelized calibration.
- Tested in supercomputer: Installation-Compilation is not an easy process. Asking for module installations, dependencies, etc.
- 6 months modelling for high resolution takes around 72 hours (Only hydrodynamic). Water Quality up to 3 hours.
- Output: up to few GB depending on the resolution.

Infrastructure: EGI

- EGI (European Grid Initiative, www.egi.eu) is a consortium of institutions that provides computing resources to different research communities.
- Researchers from all disciplines have easy, integrated and open access to the advanced digital capabilities, resources and expertise needed to collaborate and to carry out data/compute intensive science and innovation.
- Mission: Create and deliver open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries.

Infrastructure: EGI FedCloud

- Federation of EGI sites that provide cloud resources.
- What is cloud?
  - Remote Computing/Storage
  - Different paradigms: IaaS, PaaS, SaaS...
  - Transparent for the user.

EGI Federated Cloud is based on:

- Standards and validation: federation is based on common Open-Standards – OCCI, CDMI, OVF, GLUE, etc...
- Heterogeneous implementation: no mandate on the cloud technology, the only condition is to expose the chosen interfaces and services.

GEO CLARIN LifeWatch SCI-BUS Science Gateway ATLAS EXPERIMENT ENVRI DARAC VERCE esa we-nmr

Infrastructure: EGI FedCloud

- Cloud is an easier than supercomputer environment in terms of installation, compilation, calibration.
- EGI FedCloud provides enough resources for using Delft3D in high resolution.
- Attaching public IP allows user to access and get results easily. Graphic output checking allowed.
- Tested in two sites:
  - IFCA: Ubuntu 12.10, 8 VCPUs, 14G memory, 190G disk. Windows 7. D3D image.
  - CESNET: Ubuntu 12.04, 4 VCPUs, 15G memory, 10G of disk.
- Steps: certificate, VO, select site, select resources.
- Performance Cloud vs. HPC in this case depends on the type of CPU, but is similar in the tested cases.

Content

1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

Use case: Overall impact

- Eutrophication impacts directly in water quality and human health.
- Reproduce the process in other reservoirs with the same problem.
- Main reasons for eutrophication to happen.
- Develop/integrate tools for eutrophication management optimization.
- Find/Deploy the best infrastructure. Good example as Use Case.
- Needs of the project: computing power, storage, distributed workspace, etc.
- Very good Use Case for Infrastructure deployment.

Use case: Lifewatch

- **LifeWatch** is the European e-Science infrastructure for biodiversity and ecosystem research. ESFRI
- Aims to provide advanced capabilities for research on the complex biodiversity system in terms of infrastructure: computing, storage, virtual research environments, etc.
- This Use Case has been used as example for covering the data lifecycle:
  - Data acquisition system
  - Data Management Planning
  - Preservation of complete data lifecycle, including analysis and reproducibility (D3D).

Use case: INDIGO-DataCloud

- **INDIGO-DataCloud** (Integrating Distributed data Infrastructures for Global Exploitation) aims at developing a data/computing platform targeting scientific communities, deployable on multiple hardware and provisioned over hybrid (private or public) e-infrastructures.
- INDIGO will provide resources from different sources (EUDAT, EGI, PRACE) in a centralized and easy way, including single authentication for all resources.
- This Use Case has been used because it is very representative in terms of data cycle.
- A SaaS solution would be very useful (web interface?).

INDIGO - DataCloud

Cuerda del Pozo

Server

Delft3D

ecohydros
ecología avanzada del agua

IFCA
Instituto de Física de Cantabria

Config files
Input set up
Output Analysis

Replica

FedCloud

Altamira

SugarSync

Output Analysis
Output Validation

Content

1. Introduction
2. Hydrodynamic modelling
3. Water Quality modelling
4. Infrastructure
5. Use Case
6. Conclusions

Conclusions

- Water modelling is not an easy issue: installation, calibration, number of inputs, etc.
- Eutrophication is a very complex process: a number of parameters are involved. Water Quality modelling problems need to be solved.
- It needs high resolution models -> Powerful resources.
- Cloud provides resources that SMEs cannot afford.
- There are available resources that SMEs or public institutions can use: EGI, Lifewatch, INDIGO.

Thanks for your attention

Fernando Aguilar
aguilarf@ifca.unican.es
Instituto de Física de Cantabria (IFCA)
Santander - Spain