Scalable Interactive Platform for Geographic Evaluation of Sea-Level Rise Impact Combining High-Performance Computing and WebGIS Client

Agnès Tellez-Arenas, Robin Quique, Faïza Boulahya, Gonéri Le Cozannet, François Paris, Sylvestre Leroy, Fabrice Dupros, François Robida

HAL Id: hal-02137523

https://brgm.hal.science/hal-02137523

Submitted on 23 May 2019

Scalable Spatial Data Architecture for Coastal Climate Services

Agnès Tellez-Arenas, Robin Quique, Faïza Boulahya, Gonéri Le Cozannet, François Paris, Sylvestre Leroy, Fabrice Dupros, and François Robida

Abstract As climate is changing, more applied information on its impacts is required to inform adaptation planning. It is a fact that during the last decade, the amount of information relevant for climate change impact assessment has grown drastically. This can be particularly illustrated in coastal areas, where a most important recent development has been the delivery of precise and accurate topography obtained by LiDAR at regional to national scales. However, these developments have not led to easier assessment of coastal climate change impacts. This is due to both to the complexity of coastal models that also depend on local natural changes and anthropogenic actions and to the difficulty to actually use such large and complex datasets. In this paper, we describe a prototype of web service to quickly communicate spatial information on future flooding along the French coastal zones. We discuss several issues related to data architecture at large scale, on-the-fly (geo)-processing capabilities, management of asynchronous workflows and data diffusion strategies in the context of international standards such as INSPIRE (Infrastructure for Spatial Information in Europe). We believe that our flexible architecture, mainly reusing off-the-shelf components is able to improve both complex scenarios analysis for experts and dissemination of these future coastal changes to the general public.

1 Introduction

Climate change is a most important challenge for coastal zones over the 21st century and beyond. Continuous observations of coastal and marine systems have highlighted its impacts to the marine environment, the biodiversity and sea-level rise, which is a key concern for coastal risks such as shoreline changes and coastal flooding [8]. While sea-level has been mostly stable at least during the last two mil-

———————————

BRGM, Orléans, France

e-mail: {a.tellez-arenas,r.quique,f.boulahya,g.lecozanet,f.paris,s.leroy,f.dupros,f.robida}@brgm.fr

lennia, climate warming induced by anthropogenic greenhouse gas emissions have driven the Earth system into a second phase of deglaciation caused by the melting of glaciers and the two remaining polar ice-sheets of Greenland and Antarctica [5]. Hence, the rate of sea-level rise is currently limited to 3 millimeters per year [4], but it is expected to rise as much as 1 m to 3 m per century over the next millennia. To adapt to the adverse impacts of sea-level rise and climate change, coastal stakeholders will need validated information regarding past, current and future coastal changes. This information will be provided through climate services, one component of which being portals, providing access to heterogeneous data and processing tools validated by the scientific community [3]. In the case of coastal zones, the baseline data are mean and extreme sea-levels and the local coastal topography and lithology [9]. Besides detailed modeling tools that require significant efforts [14], methods exist to combine and process this data and provide information on the coastal vulnerability together with its uncertainties [6]. However, it is a fact that portals tailored to the needs of coastal stakeholders are currently lacking, especially in Europe , thus preventing coastal stakeholders and their traditional coastal service providers to engage in a dynamics toward the uptake of coastal services for adaptation.

Indeed, computer-aided decision support systems remain a major tool for quick assessment of natural hazards and climate change. The ability to process the amount of information from Earth observation systems at different scales is a key for the reliability of the numerical models and for the interactivity of the web portal. Although research infrastructures such as Copernicus or EPOS provide resources or services to tackle datasets at global scale, the ability to design web-based geographical information system (WebGIS) at regional scale is of great interest. Such systems showing seal-level scenarios should be built with open source components to ensure future integration in larger systems.

This paper describes a scalable architecture using a high performance computing system as a cornerstone. We discuss several issues such as large scale data management, scientific workflows and the integration of processing capabilities. The paper proceeds as follows. Section II discusses the technical background related to the main component of our architecture. Then, Section III presents an overview of our implementation and discusses the challenges we must overcome when dealing with the integration of processing facilities in a web-based workflow. Section IV illustrates the main features of our client and underlines our prefetching mechanism for the management of the tiles. Section V describes related work and Section VI concludes this paper.

2 Technical background

We briefly present in this section the technologies and standards used for the implementation of the prototype.

2.1 Web Services

Web services are components publishing processes and data sets through the internet. We can distinguish two main categories. Firstly, the synchronous static web services that always publish the same pre-computed data. They are requested with some parameters such as the area of interest and the required output format. The response is an image or the data in a standardized format. The second category of web services is dynamic and usually asynchronous. Each request to such a web service launches a process which can be then requested for intermediate results or to get the status of execution. At the end, the result is a dataset, published as an image of any other standard format.

2.2 Geoprocessing orchestration

If standards web services are the final products which can be then registered in official registries and catalogues, for end-users we have made available a web client displaying the pre-computed data sets, and allowing interactions with the process, depending on parameters, scale and area of interest. The orchestration tool is a cornerstone to handle these interactions and it should meet several requirements:

- handle the orchestration of the request to the Web Services;
- manage both synchronous and asynchronous requests to the computing architecture;
- consume intermediate results and handle possible fails and delays in the process;
- ability to cancel the process during the computation.

We have tested several scientific orchestration tools and have implemented Taverna [10]. This open source software includes several connectors and a graphical interface to build the workflow. Taverna also offers a server to run the processes. However each request through Taverna takes a few seconds to launch the server. It is therefore only suitable for asynchronous usage.

2.3 High Performance Computing

The needs for large scale processing capabilities have been widely recognized in order to shift from static to dynamic spatial data architecture. Indeed, the challenge of delivering near-real-time information at regional scale relies on the availability and the efficient exploitation of parallel processing resources. Data transfer between the dedicated storage attached to the computing system (*i.e.*high performance parallel file system) and the heterogeneous and virtual storage space associated with the web components (*i.e.*virtual machines accessing classical networked storage) are, therefore, critical. One needs to keep the data as close as possible to the processing

facilities regarding our distributed system.

Efficient scheduling of the computing tasks represents another issue. The batch scheduler is both in charge of the correct usage of the physical resources available but must also provide the information to the Web-based client in terms of job status, execution time or failure.

2.4 On top of standards

In the domain of web services for georeferenced datasets, the standards are defined by Open Geospatial Consortium (OGC)[1] , and we are implementing:

- WMS: Web Map Service: synchronous static services delivering maps as images;
- WFS: Web Feature Services: synchronous static services delivering maps in a standardized text format;
- WPS: Web Processing Services: (a)synchronous services executing a process, allowing on-the-fly computing.

Implementation of OGC standards allows to register web services in international registries such as GEOSS (Global Earth Observation System of Systems). We also plan to publish Web Services in the European initiative INSPIRE (Infrastructure for Spatial Information in Europe) such as a WFS delivering pre-computed data about flooding risk computed for population areas and land cover, according to the INSPIRE official data specification on Natural Risk Zone. We use open source software to implement standard OGC web services: MapServer (WMS 1.3.0 and WFS 2.0) and GeoServer (WPS 1.0.0)

3 Implementation

3.1 Architecture overview

We have designed our scalable WebGIS architecture on the shared foundation describes in section 2. The challenge is to propose a user-friendly interface to end-users with good performances and interactivity while computations are running on big datasets. The management of the complex workflows involved represents another challenge [12]. In order to fulfill those requirements, we have implemented a distributed architecture composed by:

- A web server running the web application (Apache, Tomcat, Java): implementation of the web clients;
- web services (GeoServer Web Processing Service, according to the standard of the Open Geospatial Consortium);

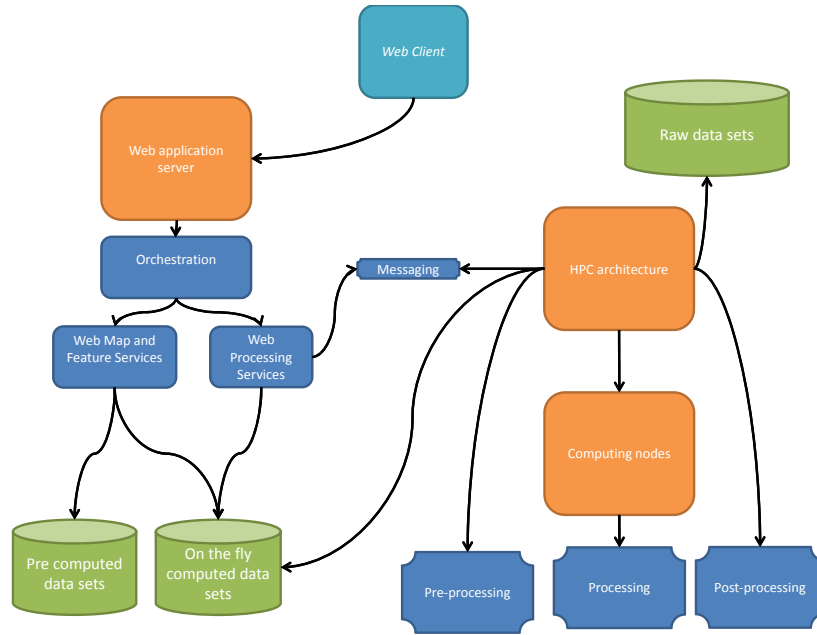
[1] http://www.opengeospatial.org/

Fig. 1: Overview of our coastal climate change information system that combines GIS, web services and processing capabilities.

- messaging components implementing secured communications between web services and the processing architecture;
- the High Performance Computing facilities dedicated to the fast processing of the data.

The detailed workflow is illustrated in Figure 1 and 2. We consider the use of a web processing service (implementation of the WPS 1.0.0 from OGC standard) that allows launching synchronous and a synchronous process with parameters, then to get the results or intermediate results together with progress information. The WPS calls HPC components then publishes the results using WMS and WFS. WMS and WFS are also used to publish pre-computed data sets. In this framework, we have developed specific process using the open source product GeoServer. Among all the standards operations implemented, GeoServer proposes a *dismiss* request allowing sending a command to stop the ongoing computation (in the case of asynchronous process).

In our case a single asynchronous WPS has been implemented. This web service has two parameters: the extent of the area of interest (bbox) and the current scale. The processing components are implemented in Python and exploits GDAL library to deal with GIS formats. All datasets are shared between the web application server, the computing architecture and the map server through a networked storage. During

the computation a dedicated parallel file system (Lustre) is used in order to maximize the I/O throughput speed.
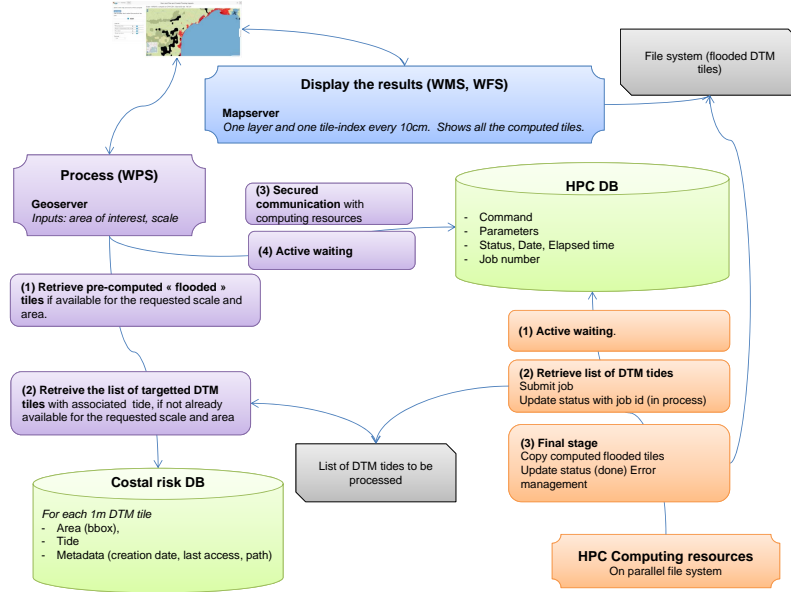

Fig. 2: Technical description of our architecture with the interactions between the different components (databases, web services, jobs scheduler, messaging tools, . . .)

3.2 High performance workflow

If we zoom in on the processing component in Figure 1, we can observe the complex workflow involved at this level. Table 1 shows the characteristics of the computing nodes. This medium scale cluster is representative of the type of resources available at the research group level.

These computing resources are shared between several users leading to particular

Peak performance	# computing nodes	# cores per node	Network	File System	Storage
10 Teraflops	30	24	Infiniband	Lustre	15 Terabytes

Table 1: Detailed characteristics of the High Performance Computing architecture.

attention paid to meet the requirements of heterogeneous workloads. We use OAR [2] job scheduler to distribute the requests of computation over the multicore nodes. The specific computations corresponding to our web-based information system are scheduled to a dedicated queue with a highest priority. According to the available resources, the estimated execution time is sent to the web application. Otherwise, the WPS server is notified that the computing request is pending and a warning message is provided to the end-user.

Usually HPC resources are internal and must be very secured, while web applications are in DMZ area accessible from the internet (Demilitarized Zone, network area protected and separated from internal network by Firewall which filters or forbids network connections). Communication from DMZ to internal secured resources is never allowed to protect the main infrastructures and databases. In our architecture, the secured communication between those two areas is based on a messaging components (RabbitMQ [3]) implementing *one direction only* communications between web services and HPC components: Web components put messages in a secured queue. Internal HPC component reads the messages. There is not any connection from the internet area to the internal network but only from internal network to DMZ.

So in our application the WPS server submits two kinds of requests : the submission and the cancellation of a calculation. The submission of a calculation means that the WPS server puts a request message in the dedicated RabbitMQ queue. The HPC cluster gets this request, copies all the input data (DTM tiles) to the Lustre file system and submits the flooding job to the dedicated OAR queue. According to the number of tiles to process, the resolution and the available computing resources, the HPC cluster pushes in the RabbitMQ queue a message describing the job (identifier, estimated time, start date, status) in order the WPS server knows what to do. During each step of calculation those information are updated until the end of the job. Figure 3 summarizes this workflow. As tiles are numerous and independent, several groups of tiles are computed in parallel. This strategy allows us to fully benefit from the underlying computing power available and significantly speed up the overall workflow even for fine resolution scenarios. In each group, for each tile, the calculation consists in three steps : (1) convert in a new coordinate system, (2) recover the corresponding Highest Astronomical Tide (HAT) and (3) finally compute the flooding for ten levels from 10 cm to 100 cm.

The cost of the copy of the output data of the current tile from the high performance storage to the global shared storage is overlapped by the computation of the next tile. As the final WMS and WFS are efficiently dealing with the tiles, we do not merge the tiles : the web client can display intermediate results. During his session an end-user can try to cancel a flooding job. In this case, the WPS server post a message to the RabbitMQ queue, that is read by the HPC cluster that cancels the running job.
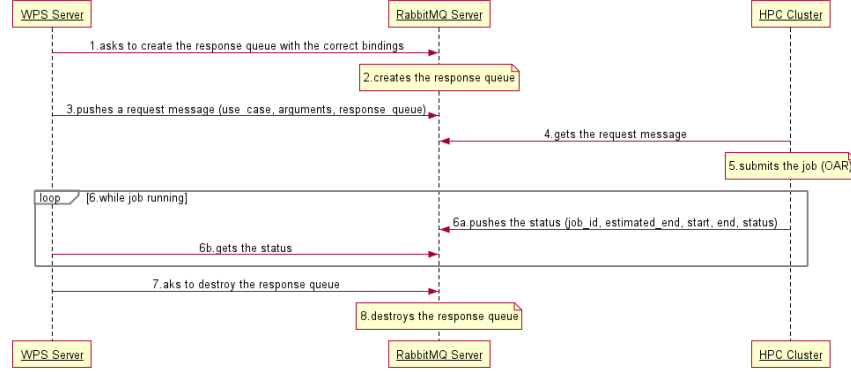
---

[2] https://oar.imag.fr/

[3] https://www.rabbitmq.com/

Fig. 3: Flowchart of the submission of a job scheduled on the computing cluster platform.

4 Illustration of usage

4.1 WebGIS client

The French coastlines display contrasted geomorphological features which all are differently vulnerable to sea-level rise [13]. Concerns are especially high in south-western and Mediterranean coasts, which display low-lying areas and erodible beaches and wetlands, leading to the adaptation guidelines, which must be applied at local to regional scales and encourage climate-compliant land use practices [11]. Compared to other countries such as the UK, France has acquired high resolution topographic data (LiDAR) only recently. These datasets are now available and can be combined with the reference mean and extreme water levels (provided by the SHOM) in order to identify areas prone to flooding.

Our architecture combines web approach, GIS-functionality and capabilities to process large datasets (Figure 4 and 5) Functionalities are accessible to a user through a web portal, proposing an interactive interface with usual map viewer tools (zoom in, zoom out, hide or show layers,...). The user selects the area of interest and the web browser automatically displays pre-computed datasets is available, depending on the scale. Pre-computed flooded tides are available for a scale greater than 1:50k . If the user zooms in to a detailed scale for which no tides are already computed, he can launches the process then the new flooded tides are automatically displayed as soon as computed. At the scale, a more detailed DMT is used for the computation. The user can also easily choose the level of see rising (10 cm, 20 cm,...,1 m) we wants to display: this does not requires a new computation, every level being computed in the same process.
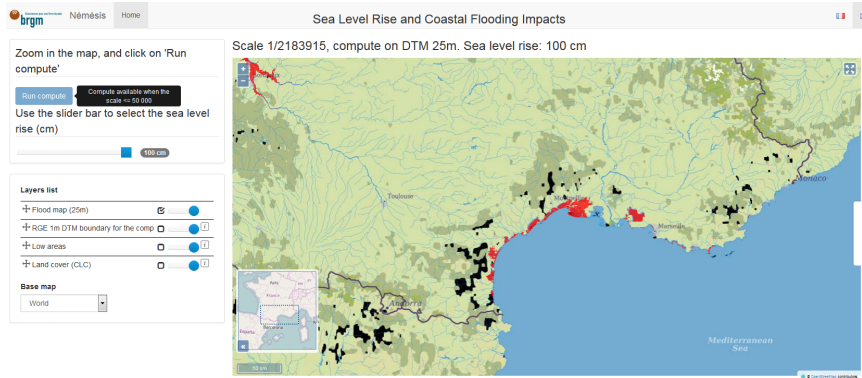
Fig. 4: WebGIS client based on the 25 m-DTM with a scenario of sea level rise at 100 cm.
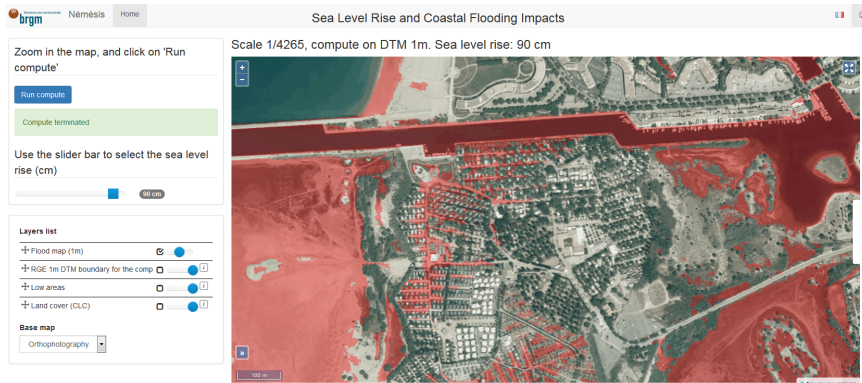

Fig. 5: WebGIS client based on the 1 m-DTM with a scenario of sea level rise at 90 cm. Red color shows the flooded area.

4.2 Tiles management

Our one-meter DTM covers all the French costal area and is composed of a huge amount of tiles (each file is of size 20 MB) covering an area of one square kilometer. To deal with different scales, we use this one-meter DTM (local scale) to derive a 10-meter DTM (regional scale) and 25-meter DTM (national scale). At national scale (scale greater than 1:50K), the end-user visualizes pre-computed flooding of the French coast for a level rise of 50 cm and 100 cm. For regional scale (greater than 1:5000), flooding is computed on-the-fly on the 10-meter DTM. In the same way, for local scale (less than 1:5000), flooding is computed on the fly on the one-meter DTM.

While computation is running, intermediate results are showed to the end-user in

such a way that he can start to explore and visualize. The user will also be warned of the waiting time before getting the final result. In case of real-time computations, the web client shows intermediate results of the simulation at a frequency as close as possible of the simulated time. Pre-computed datasets are used at this level when the corresponding tiles are available.

5 Related Work

Regarding spatial data infrastructures, several wingspan initiatives have been dedicated to the deployment of robust architectures in various domains (i.e transportation, defense, space) . This is particularly true for the geosciences community where the spatial component plays a significant role. At the European level, research infrastructures like Copernicus (Earth observation) or EPOS (European Plate Observing System) provide resources and services to the associated scientific community. Most of these platforms rely on geospatial standards from the OGC ensuring interoperability and reuse across data and the associated services. At global scale, initiatives like GEOSS, One Geology represent a long term effort for research data architectures in geosciences.

The trend to enhance sharing of research data is also materialized by several organizations or initiatives promoting working groups, tools or architectures in this direction. For instance the RDA (Research Data Alliance) and the EUDAT (European Data infrastructure) are very active in terms of community building.

In the context of climate change and the impact of rising sea levels, coastal areas are a key issue for risk prevention policies. Several online coastal information system have been set up. For instance, the information system managed the US National Oceanic and Atmospheric Administration (NOAA) is focused on sea level rise and coastal flooding impacts for the national territory [1]. The viewer provides access to several layers including socio-economic data and significant efforts have been devoted to public dissemination. This information system is designed in a static way meaning that all the tiles available are pre-computed.

In the same area, the Australian coastal information system provides information on sea level rise [2]. Some pre-computed maps could be downloaded (in image format) depending on the region of interest. Introducing flexibility and scalability in such information systems remains a challenge.

6 Conclusions and Perspectives

This paper discusses a scalable spatial data architecture devoted to coastal climate services. Mainly reusing off-the-shelf components, we have designed a dynamic information system able to deliver accurate sea water elevation at various scales. Efforts to build our system on top of OGC standards are a guarantee to maintain

interoperability between our architecture and other initiatives.

Lying at the heart of our web-based GIS, the exploitation of a supercomputing architecture provides scalability and near real-time processing for the user. The efficiency of this strategy strongly depends on a fine control over both the scheduling of the processing tasks and the locality of the data.

In order to extend the capabilities of our architecture, our ongoing efforts are twofold. Firstly, the processing capabilities of our system could be extend to take into account the growing amount of environmental data available. For instance, we plan to consider physically-based modelling tools, including uncertainty quantification strategies, in order to improve our system [7]. Secondly, we plan to extend the data set available (economy, urban planning, geology). This point is of great interest for decision makers. Thus, the long term effort is to shift toward predictive analytics for coastal climate change [15].

References

1. https://coast.noaa.gov/slr/.
2. http://www.ozcoasts.gov.au/.
3. BRASSEUR, G. P., AND GALLARDO, L. Climate services: Lessons learned and future prospects. *Earth's Future 4*, 3 (2016), 79–89.
4. CAZENAVE, A., AND COZANNET, G. L. Sea level rise and its coastal impacts. *Earth's Future 2*, 2 (2014), 15–34.
5. CLARK, P. U., SHAKUN, J. D., MARCOTT, S. A., MIX, A. C., EBY, M., KULP, S., LEVERMANN, A., MILNE, G. A., PFISTER, P. L., SANTER, B. D., ET AL. Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nature Climate Change* (2016).
6. COZANNET, G. L., GARCIN, M., BULTEAU, T., MIRGON, C., YATES, M., MÉNDEZ, M., BAILLS, A., IDIER, D., AND OLIVEROS, C. An AHP-derived method for mapping the physical vulnerability of coastal areas at regional scales. *Natural Hazards and Earth System Sciences 13*, 5 (2013), 1209–1227.
7. COZANNET, G. L., ROHMER, J., CAZENAVE, A., IDIER, D., VAN DE WAL, R., DE WINTER, R., PEDREROS, R., BALOUIN, Y., VINCHON, C., AND OLIVEROS, C. Evaluating uncertainties of future marine flooding occurrence as sea-level rises. *Environmental Modelling & Software 73* (2015), 44 – 56.
8. GATTUSO, J.-P., MAGNAN, A., BILLÉ, R., CHEUNG, W., HOWES, E., JOOS, F., ALLEMAND, D., BOPP, L., COOLEY, S., EAKIN, C., ET AL. Contrasting futures for ocean and society from different anthropogenic co2 emissions scenarios. *Science 349*, 6243 (2015), aac4722.
9. GORNITZ, V. Global coastal hazards from future sea level rise. *Palaeogeography, Palaeoclimatology, Palaeoecology 89*, 4 (1991), 379–398.
10. HORVÁTH, F., ITTZÉS, P., ITTZÉS, D., BARCZA, Z., DOBOR, L., HIDY, D., MAROSI, A., AND HARDISTY, A. *Supporting environmental modelling with Taverna workflows, web services and desktop grid technology*, vol. 1. International Environmental Modelling and Software Society, 2014, pp. 423–430.
11. LE COZANNET, G., AIT-KACI, A., COLAS, S., DE LACAZE, X., LECACHEUX, S., MIRGON, C., PEINTURIER, C., GARCIN, M., AND OLIVEROS, C. Recent GIS based national assessments of climate change consequences in france: methods, results and lessons learnt. *Journal of Coastal Research 65*, sp2 (2013), 1421–1426.

12. MATTOSO, M., DIAS, J., NA, K. A. O., OGASAWARA, E., COSTA, F., HORTA, F., SILVA, V., AND DE OLIVEIRA, D. Dynamic steering of HPC scientific workflows: A survey. *Future Generation Computer Systems 46* (2015), 100 – 113.

13. PASKOFF, R. P. Potential implications of sea-level rise for france. *Journal of Coastal Research 20*, 2 (2004), 424–434.

14. ROY, S. L., PEDREROS, R., ANDRÉ, C., PARIS, F., LECACHEUX, S., MARCHE, F., AND VINCHON, C. Coastal flooding of urban areas by overtopping: dynamic modelling application to the johanna storm (2008) in gâvres (france). *Natural Hazards and Earth System Sciences 15*, 11 (2015), 2497–2510.

15. SCHNASE, J. L., DUFFY, D. Q., TAMKIN, G. S., NADEAU, D., THOMPSON, J. H., GRIEG, C. M., MCINERNEY, M. A., AND WEBSTER, W. P. MERRA analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. *Computers, Environment and Urban Systems* (2014), –.