TECHNICAL UNIVERSITY OF CLUJ-NAPOCA ROMANIA

UNIVERSITÉ DE GENÈVE

Faculty of Automation and Computer Science

Faculty of Sciences
Earth and Environmental Sciences
Institute for Environmental Sciences

Efficient Development and Execution of Environmental Applications on High Performance Parallel and Distributed Computing Infrastructures

PhD Thesis
Short Summary

Eng. Diana Denisa RODILĂ

Scientific Supervision,
Prof. Eng. Dorian GORGAN, Ph.D.
Senior Lecturer Nicolas RAY, Ph.D.
Prof. Anthony LEHMANN, Ph.D.

- 2016 –

SUMMARY

General Presentation of the Research Theme

At the beginning of the 21st century, global changes linked to climate, biodiversity and habitat loss, environmental degradation and pollution, are threatening our natural environment and the human society at large, with already tangible negative outcomes. Intensified droughts, ocean acidification, global sea level rise, increases in frequency of extreme weather events and glaciers melting are examples of such outcomes that are thought to intensify if appropriate international policies are not endorsed and applied. Responding effectively to all these complex changes has become an important challenge for policy makers, but also for the scientific community that demands access to continuously increasing quantities of heterogeneous data and resources. The need to understand the inter-linkage between natural phenomena and human-induced activities is urgent and an important aspect for achieving this is the accessibility and processing of environmental data from various disciplines and geographic scales (local, regional, national and global).

Scientists started to look and feel the need of an "innovative solution" which will allow them to access an operational infrastructure, supporting large scale multidisciplinary applications while providing highly elastic resources. The computational and storage capacity required for such a challenge exceeds most of the time what an average computational center can offer.

In this thesis we propose a general methodology and framework for easily porting and executing environmental applications simultaneously on different parallel and distributed infrastructures such as cluster, Grid and Cloud. To achieve this goal, we focus on solving two main problems: 1) the interoperability and the coexistence of different distributed, heterogeneous computing infrastructures within a Hybrid Computing Environment, and 2) the interoperability between Environmental Sciences (environmental data and environmental applications) and such a Hybrid Computing Environment and its underlyinf infrastructures.

The methodology proposed in this research will allow users to interact in a central and transparent manner with several parallel and distributed infrastructures to try to solve the most challenging issues that are threatening our environment.

Research Questions

The main goal of this research is: **To analyze and explore in which way Information Technology, and especially parallel and distributed high-performance systems, can improve the major challenges and needs the Environmental Sciences are facing in the process of extracting understandable and useful information from raw environmental data, leading in the end to a better informed decision making towards a sustainable development and a sustainable planet.**

To be able to achieve this goal, we have formulated some associated research questions that we will try so answer on the chapters of this thesis:

1. **What are the current urgent *needs and challenges* of Environmental Sciences (focusing on environmental data and environmental applications)?**
2. **What is the *state of the art* of High Performance Computing (HPC) landscape (Cluster, Grid, Cloud, Hybrid Computing)?**
3. **What are the *lessons learned* based on a heuristic approach of integrating Environmental applications with HPC?**
4. **What are the flexible solutions to *integrate* environmental applications and environmental data with HPC infrastructures?**
5. **How can we solve the *interoperability* between Environmental Sciences and Computer Sciences?**
6. **Is there an efficiently way to take advantage of all the available *heterogeneous computing* infrastructures *simultaneously*?**
7. **How can we solve the *interoperability between different HPC infrastructures*? What are the challenges and how this solution can be *applied* to Environmental Sciences?**
8. **How can we *evaluate and validate* the proposed solutions?**

Structure of the Thesis

This thesis is structured in five big sections, which are further sub-structured in twelve Chapters. This structure is illustrated in Figure 1.

- **Introduction –** is the introductory part which presents the big picture (general context) of the thesis, explaining the need for such a research work and the interdisciplinary nature of the thesis. This section contains one chapter.
    - **Chapter 1** – presents the general interdisciplinary research theme, what are the major challenges and research questions we address during the thesis and what are our main objectives for this work.
- **State of the Art –** gives a recent overview of the themes and fields approached in this thesis, both in Environmental Sciences and Computer Science research areas. This section contains four chapters:
    - **Chapter 2** – focuses on the existing data challenges present in Environmental Sciences and introduces environmental data together with its main attributes, properties and transformations.
    - **Chapter 3** – gives an overview on the basic concepts and notions underlying the Spatial Data Infrastructures, focusing on standards, interoperability, tools and initiatives used to support this idea.
    - **Chapter 4** – introduces environmental applications, their specificities and characteristics and what are the challenges related to them. This chapter focuses on two main categories of environmental applications: hydrological modeling and remote sensing applications.

- **Chapter 5** – presents a general updated view of distributed systems (Cluster, Grid, Cloud, etc.), focusing on the properties and tools that have been used more intensively in our research. We also introduce the concept of Hybrid Computing and we generally present the properties of distributed applications.

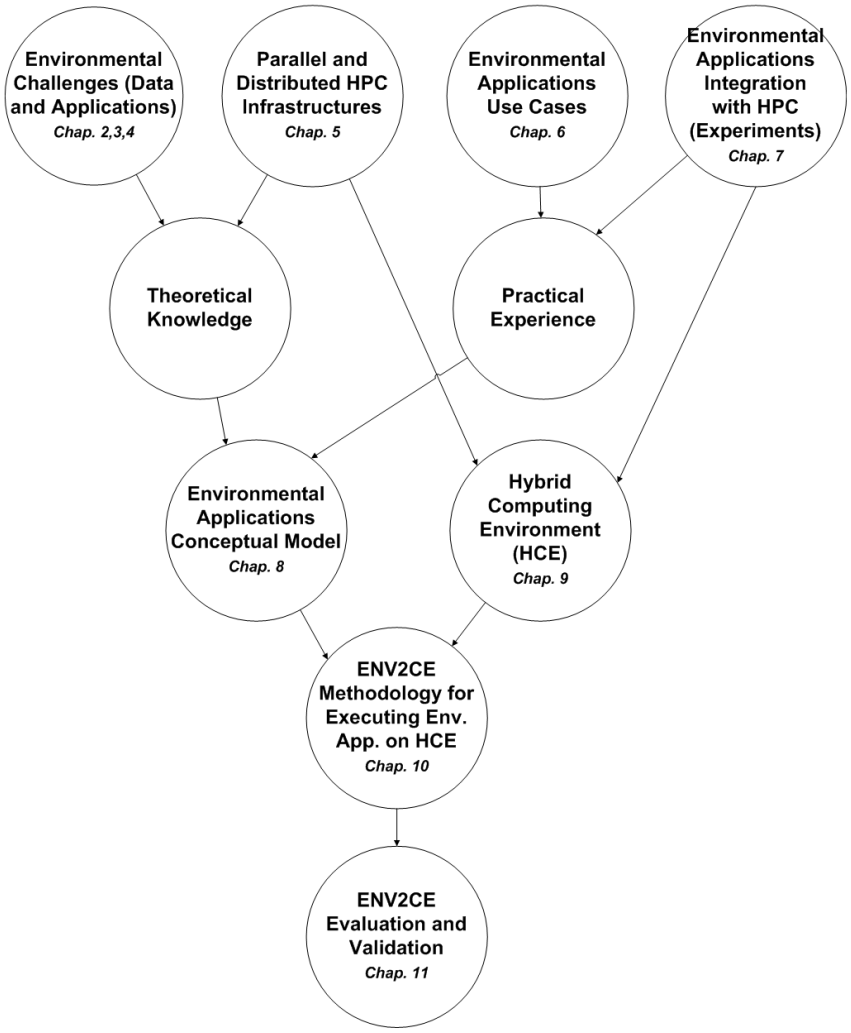

Figure 1: Thesis Structure

- **Development and Execution of Environmental Applications –** describes in detail the main environmental applications used and developed during our research. This section is composed out of two chapters:
  - **Chapter 6** – presents a set of environmental applications that were developed and/or used during our research in different national and international projects.
  - **Chapter 7** – gives an overview of the growing needs and requests identified in Environmental Sciences to take advantage of the capabilities offered by the parallel and distributed infrastructures. This chapter focuses on a list of

experiments addressed in our research of porting different environmental applications on different parallel and distributed infrastructures. We analyzed the results, drew the conclusions and came up with a list of lessons learned based on the performed experiments. These lessons learned are the foundation of the next chapters, as we based our next proposals on our gained experience so far.

- **New Methodology and Framework Proposal –** introduces a new methodology (ENV2CE) and a new framework for efficiently porting and executing evironmental applications on a Hybrid Computing Environment (HCE). This section contains four chapters:
  - **Chapter 8** – presents a conceptual description of environmental applications by introducing a proposal for an application conceptual model together with the methodology to apply this model.
  - **Chapter 9** – gives an overview of what a HCE is, what are the properties, advantages, disadvantages and most of all the challenges of working with such an environment.
  - **Chapter 10** – focuses on the new methodology and the new framework proposed for efficiently executing environmental applications on a HCE. This chapter presents the methodology phases, the system architecture and describes the main components together with their main functionalities.
  - **Chapter 11** – presents the evaluation and the validation of the proposed solutions.
- **Conclusions –** presents the general conclusions and contains one chapter:
  - **Chapter 12** – concludes this research by answering the research questions addressed in the introduction, by emphasizing our contributions and by presenting recommendations for future directions of work.

Published Articles

Journal Papers

1. **Rodila, D.,** Gorgan, D., Ray, N., and Lehmann, A. (2016). ENV2CE: Environmental Application Conceptualization and Execution on a Hybrid Computing Environment - Framework and Methodology Proposal, (to be submitted).
2. **Rodila**, **D.,** Ray, N., and Gorgan, D. (2015). Conceptual Model for Environmental Science Applications on Parallel and Distributed Infrastructures, Environmental System Research, Vol. 4/23, 2015, DOI: 10.1186/s40068-015-0050-1.
3. **Rodila**, **D.,** Bacu, V., and Gorgan, D. (2012). Comparative Parallel Execution of SWAT Hydrological Model on Multicore and Grid Architectures, in International Journal of Web and Grid Services (IJWGS), Vol. 8/3, September 2012, DOI: 10.1504/IJWGS.2012.049172, pp.304 - 320.
4. **Rodila**, **D.,** and Gorgan, D. (2012). Geospatial and Grid Interoperability through OGC Services Gridification, in International Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), Vol. 5/6, December 2012, DOI: 10.1109/JSTARS.2012.2217115, ISSN: 1939-1404, pp. 1650 - 1658.
5. Giuliani, G., Dao, H., De Bono, A., Chatenoux, B., Allenbach, K., De Laborie, P., **Rodila, D.,** Alexandris, N., and Peduzzi, P. (2016). Live Monitoring of Earth Surface (LiMES): a framework for monitoring environmental changes from Earth Observations, (to be submitted).
6. Silvestro, F., Campo, L., Rudari, R., De Angeli, S., D'Andrea, M., **Rodila**, **D.,** and Gabellani, S. (2016). Impacts of EC-Earth Global Climate Model RCP4.5 climate change scenario on maximum daily streamflow quantiles at global scale, Journal of Climate (submitted article).
7. Mihon, D., Colceriu, V., Bacu, V., Allenbach, K., **Rodila**, **D.,** Giuliani, G., and Gorgan, D. (2013). OGC Compliant Services for Remote Sensing Processing over the Grid Infrastructure, in International Journal of Advanced Computer Science and Applications (IJACSA), ISSN 2158-107X, pp. 32–40.
8. Colceriu, V., Mihon, D., Minculescu, A., Bacu, V., **Rodila**, **D.,** and Gorgan, D. (2013). Workflow Based Description and Distributed Processing of Satellite Images, in International Journal of Advanced Computer Science and Applications (IJACSA), ISSN 2158-107X, pp. 50–57.
9. Bacu, V., Mihon, D., Stefanut, T., **Rodila**, **D.,** Abbaspour, K., Rouholahnejad, E., and Gorgan, D. (2013). Calibration of SWAT Hydrological Models in a Distributed Environment Using the gSWAT Application, in International Journal of Advanced Computer Science and Applications (IJACSA), ISSN 2158-107X, pp. 66–74.
10. Gorgan, D., Bacu, V., Stefanut, T., **Rodila**, **D.,** and Mihon, D. (2012). Earth Observation application development based on the Grid oriented ESIP satellite image processing platform, in Journal on Computer Standards & Interfaces, Published by Elsevier B.V., doi:10.1016/j.csi.2011.02.002, ISSN: 0920-5489, pp: 541–548.
11. Gorgan, D., Bacu, V., Mihon, D., Stefanut, T., **Rodila**, **D.,** Cau, P., Abbaspour, K., Giuliani, G., Ray, N., and Lehmann, A. (2012). Software platform interoperability throughout enviroGRIDS portal, in International Journal of Selected Topics in Applied Earth Observations and Remote Sensing - JSTARS, Vol. PP/99, pp. 1-11.

12. Gorgan, D., Bacu, V., Mihon, D., **Rodila**, **D.,** Abbaspour, K., and Rouholahnejad, E. (2012). Grid based calibration of SWAT hydrological models, in Journal of Nat. Hazards Earth Syst. Sci., Vol. 12/7, pp. 2411-2423, doi:10.5194/nhess-12-2411-2012.
13. Mihon, D., Bacu, V., **Rodila**, **D.,** Stefanut, T., Abbaspour, K., Rouholahnejad, E., and Gorgan, D. (2012). Grid Based Hydrologic Model Calibration and Execution, Chapter in the book: Advanced in Intelligent Control Systems and Computer Science, Dumitrache I. (Ed.), Springer-Verlag, ISBN 978-3-642-32548-9, DOI: 10.1007/978-3-642-32548-9_20, Volume 187, pp 279-293, 2012.
14. Gorgan, D., Bacu, V., **Rodila**, **D.,** Pop, F., and Petcu, D. (2010). Experiments on ESIP - Environment Oriented Satellite Data Processing Platform, in Earth Science Informatics Journal, Springer, Vol.3/4, DOI: 10.1007/s12145-010-0065-0, ISSN: 1865-0473, pp. 297-308, December 2010, http://www.springerlink.com/content/v463w6872266j282/.

Conference Papers

1. **Rodila**, **D.,** and Gorgan, D. (2012). Mapping Geospatial Applications onto Parallel and Distributed Environments, in ePaMuS 2012 – 5th International Workshop on Engineering Parallel and Multi-Core Systems – The Multi-Core Workshop, Palermo, Italy, 4-6 July, 2012, DOI: http://dx.doi.org/10.1109/CISIS.2012.152 , pp. 443 - 448.
2. **Rodila, D.,** Bacu, V., and Gorgan, D. (2012). Geospatial Applications on Different Parallel and Distributed Systems in enviroGRIDS Project, in European Geosciences Union - General Assembly EGU 2012, Vienna, Austria, April 22-27, 2012, (abstract).
3. **Rodila**, **D.,** Bacu, V., and Gorgan, D. (2011). Comparative Analysis of Distributed and Grid Based Execution of SWAT Model, in 3PGCIC 2011 - Sixth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Barcelona, Spain, October 26-28, 2011, DOI: http://dx.doi.org/10.1109/3PGCIC.2011.49, pp. 273-278.
4. **Rodila, D.,** and Gorgan, D. (2011). A Mediation Approach in Geospatial Web Services Gridification, in ICCP2011 – IEEE International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, August 25-27, 2011, DOI: http://dx.doi.org/10.1109/ICCP.2011.6047928, pp 541-548.
5. **Rodila, D.,** Bacu, V., Ardelean, V., Borlea, C., and Gorgan, D. (2011). Geospatial Web Services Gridification in enviroGRIDS, in European Geosciences Union - General Assembly EGU 2011, Vienna, Austria, April 03-08, 2011, (abstract and presentation), http://meetingorganizer.copernicus.org/EGU2011/EGU2011-11469.pdf .
6. Gorgan, D., Mihon, D., Bacu, V., **Rodila, D.,** Stefanut, T., Colceriu, V., Allenbach, K., Balcik, F., Giuliani, G., Ray, N., and Lehmann, A. (2013). Flexible Description of Earth Data Processing over HPC Architectures, Big Data From Space Symposium, Frascati, Italy, 5-7 June, Abstract Book, pp. 39.
7. Bacu, V., **Rodila, D.,** Pop, O., Dumitru, C., and Gorgan, D. (2012). Graphical Simulation and Visualization of Human Organs, MEI 2012 - International Conference on Medical Education Informatics, 6-7 April, Thessaloniki, http://www.mei2012.org/content/graphical-simulation-and-visualization-human-organs.
8. Bacu, V., Mihon, D., Stefanut, T., **Rodila, D.,** Gorgan, D., Cau, P., and Manca, S. (2011). Grid Based Services and Tools for Hydrological Model Processing and Visualization, in SYNASC 2011 – 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, September 26-28.

9. Bacu, V., Mihon, D., **Rodila, D.,** Stefanut, T., and Gorgan, D. (2011). Grid Based Architectural Components for SWAT Model Calibration, in HPCS 2011 - International Conference on High Performance Computing and Simulation, Istanbul, Turkey, ISBN 978-1-61284-381-0, pp. 193-198, July 4-8.
10. Bacu, V., Mihon, D., **Rodila, D.,** Stefanut, T., and Gorgan, D. (2011). gSWAT Platform for Grid based Hydrological Model Calibration and Execution, in ISPDC 2011 - 10th International Symposium on Parallel and Distributed Computing, Cluj-Napoca, Romania, July 6-8, 2011, pp.288-291.
11. Gorgan, D., Bacu, V., Mihon, D., Stefanut, T., **Rodila, D.,** Kokoszkiewicz, L., Rouholahnejad, E., Abbaspour, K., and van Griensven, K. (2011). Grid Based Hydrological Model Calibration and Execution by gSWAT Application, 2011 International SWAT Conference, 15-17 June, 2011, Toledo, Spain, 2011.
12. Gorgan, D., Bacu, V., Manca, S., Giuliani, G., **Rodila, D.,** Stefanut, T, Mihon, D., Abbaspour, K., Rouholahnejad, E., van Griensven, A., Kokoszkiewicz, L., and Cau, P. (2011). Tools and Applications in enviroGRIDS Project for Spatial Data Processing of Black Sea Catchment Basin, EGI User Forum, Vilnius, 11-15 April, https://www.egi.eu/indico/contributionDisplay.py?contribId=154&sessionId=16&confId=207.
13. Bacu, V., **Rodila**, **D.,** Kokoszkiewicz, L., Rouholahnejad, E., Mihon, D., van Griensven, A., Abbaspour, K., and Gorgan, D. (2011). Calibration and Execution of SWAT Models over GRID Architecture, Presentation to ESSI-12, EGU 2011, Vienna, 3-8 April, 2011.
14. Gorgan, D., Mihon, D., Stefanut, T., Bacu, V., **Rodila, D.** (2011). Grid Based Environment Oriented Tools and Applications for Black Sea Catchment Basin, Presentation to HS12.8, EGU 2011, Vienna, 3-8 April, 2011.
15. **Rodila, D.,** and Gorgan, D. (2010). Integration of Spatial Data Infrastructures with Grid Environment, in SYNASC 2010 – 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, September 23-26, 2010, DOI: http://dx.doi.org/10.1109/SYNASC.2010.63, pp. 269-277.
16. **Rodila, D.,** Gorgan, D., and Bacu, V. (2010). The Interoperability between OGC Services and Grid Environment in EnviroGRIDS Project, in 3PGCIC 2010 – International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Fukuoka, Japan, November 4-6, 2010, IEEE Computer Press, ISBN: 978-0-7695-4237-9, DOI: http://dx.doi.org/10.1109/3PGCIC.2010.65, pp. 387-392.
17. Gorgan, D., Stefanut, T., Bacu, V., Mihon, D., and **Rodila, D.** (2010). Grid based Environment Application Development Methodology, in Large-Scale Scientific Computing, Springer Journal LNCS 5910, ISBN 978-3-642-12534-8, pp.499-506, DOI: 10.1007/978-3-642-12535-5_59.
18. Gorgan, D., **Rodila**, **D.,** Bacu, V., Giuliani, G., and Ray, N. (2010). OGC and Grid Interoperability in enviroGRIDS Project, in European Geosciences Union - General Assembly EGU 2010, May 02-07, 2010, Vienna, Austria (abstract and presentation), http://meetingorganizer.copernicus.org/EGU2010/EGU2010-13457.pdf .
19. Gorgan, D., **Rodila, D.,** Bacu, V., Giuliani, G., Ray, N., Charvat, K., and Lehman, A. (2010). Geospatial and Grid infrastructures interoperability in enviroGRIDS, 5th EGEE User Forum, April 12-16, 2010, Uppsala, Sweden (2010). Book of Abstracts, http://indico.cern.ch/contributionDisplay.py?contribId=161&confId=69338.

20. Bacu V., Gorgan, V., **Rodila, D.,** Pop, F., Neagu, G., and Petcu, D. (2010). gProcess and ESIP Platforms for Satellite Imagery Processing over the Grid, EGU2010 - European Geosciences Union General Assembly, 02 – 07 May 2010, Vienna, Austria (2010). Geophysical Research Abstracts, Vol. 12, EGU2010-14125-1, eISSN: 1607-7962 (2010). http://meetingorganizer.copernicus.org/EGU2010/EGU2010-14125-1.pdf
21. Bacu, V., **Rodila, D.,** Mihon, D., Stefanut, T. and Gorgan, D. (2010). Error prevention and recovery mechanisms in the ESIP platform, in ICCP2010 – IEEE 6th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, August 26-28, 2010, pp. 411 – 417.
22. **Rodila, D.,** Bacu, V., and Gorgan, D. (2009) Semantic Annotation based Service Composition for Grid Workflow Description and Execution, in SYNASC 2009 – 11th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, September 26-29, 2009, ISBN: 978-0-7695-3964-5, DOI: http://dx.doi.org/10.1109/SYNASC.2009.11, pp.245-253.
23. **Rodila, D.,** Bacu, V., and Gorgan, D. (2009). Integration of Satellite Image Operators as Workflows in the gProcess Application, in Proceedings of ICCP2009 - IEEE 5th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, August 27-29, 2009, ISBN: 978-1-4244-5007-7, DOI: http://dx.doi.org/10.1109/ICCP.2009.5284737, pp. 355-358.
24. Gorgan, D., Bacu, V., **Rodila, D.,** Pop, F., and Petcu, D. (2009). Experiments on ESIP - Environment oriented Satellite Data Processing Platform. SEE-GRID-SCI User Forum, Bogazici University, Istanbul, Turkey, December 9-10, ISBN: 978-975-403-510-0, pp. 157-166.
25. Gorgan, D., Bacu, V., Stefanut, T., **Rodila, D.,** and Mihon, D. (2009). Grid based Satellite Image Processing Platform for Earth Observation Applications Development, IDAACS'2009 - IEEE Fifth International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 21-23 September, Cosenza, Italy, IEEE Published in Computer Press, ISBN 978-1-4244-4901-9, 247-252.
26. Bacu V., Stefanut, T., **Rodila, D.,** and Gorgan, D. (2009). Process Description Graph Composition by gProcess Platform, HiPerGRID - 3rd International Workshop on High Performance Grid Middleware, 28 May, Bucharest. Proceedings of CSCS-17 Conference, Vol.2., ISSN 2066-4451, pp. 423-430.
27. Gorgan, D., Bacu, V., Stefanut, V., **Rodila, D.,** and Petcu, D. (2009). Grid based Satellite Data Processing Platform in Earth Observation, 4th GRID & e-Collaboration Workshop – Digital Repositories, 25-26 Feb 2009, ESA/ESRIN Frascati, Italy, http://www.congrex.nl/08M16/.

Technical Notes

1. **Rodila, D.,** Chatenoux, B., and Giuliani, G. (2016). Landsat 8 resources - a short comparison of different data access providers and methods (submitted paper).