Glacier Modelling on the Cloud: Research and Education in a Web Browser

Gletschermodellierung in der Cloud: Forschung und Lehre in einem Webbrowser

http://edu.oggm.org

A project submitted to the "*Förderkreis 1669, Digital Innovation in Research and Teaching 2019*" by

Fabien Maussion, Department of Atmospheric and Cryospheric Sciences (ACINN)

fabien.maussion@uibk.ac.at

1 Context and Motivation

1.1 Global glacier modelling at University of Innsbruck

Glaciers form prominent features of many landscapes, and their global shrinking has become an icon of climate change. Glaciers are a source of geohazards, are important regulators of water availability for downstream populations and are a major contributor to sea-level rise.

There are more than 200.000 mountain glaciers in the world. Before the release of the first global inventory of glacier outlines in 2012 (the Randolph Glacier Inventory, Pfeffer et al., 2014), estimates of global glacier evolution were based on extrapolated observations and scaling relationships. Since 2012, a new branch of glaciology ("global scale glaciology") emerged to (i) estimate the volume of all glaciers worldwide (Huss and Farinotti, 2012; Farinotti et al., 2019) and (ii) their past and future evolution (e.g. Radić and Hock, 2011, 2014; Giesen and Oerlemans, 2012, 2013; Huss and Hock, 2015).

The university of Innsbruck is a leading institution in the field of global scale glaciology thanks to pioneering publications by Ben Marzeion, Georg Kaser and colleagues: Kaser et al. (2006, 2010); Marzeion et al. (2012a,b, 2014a,b, 2018). Today, the LFU hosts the development of the first global scale, open-source model of glacier evolution including ice dynamics: the Open Global Glacier Model (oggm.org, Maussion et al., 2018).

OGGM is a community model, which means that any interested research group can add or enhance parts to the model. It is the only global scale model which is open-source, and has therefore attracted new users and contributors from external institutions. It forms the basis of several research proposals and research publications originating from international teams in Germany, France, Canada, Belgium, to name only a few.

1.2 Interactive programming for research and education

Programming has become a crucial tool for all fields of science, from economics to quantum mechanics. Our students now have to learn at least one programming language on top of the "core disciplines": for example, in the master curriculum of Atmospheric Sciences taught at ACINN the "Scientific Programming" lecture yields as many ECTS as "Mountain Meteorology". A large part of the activities of modern research is to parse, explore and analyse large datasets using a programming environment.

In recent years, innovative tools have emerged to ease both the teaching of programming languages and their use for interactive data exploration. A prominent example is the "Jupyter Notebook" (jupyter.org, from now on shortened to "notebook"). Notebooks are documents that contain live computer code, equations, visualizations and narrative text, all in the same document (see Fig. 1). I have been using notebooks for research and teaching since 2015. They are extremely handy to explain new concepts to students, to grade their homework and to share information in the form of readable, reproducible computer code. We also use notebooks to write tutorials about how to use OGGM.

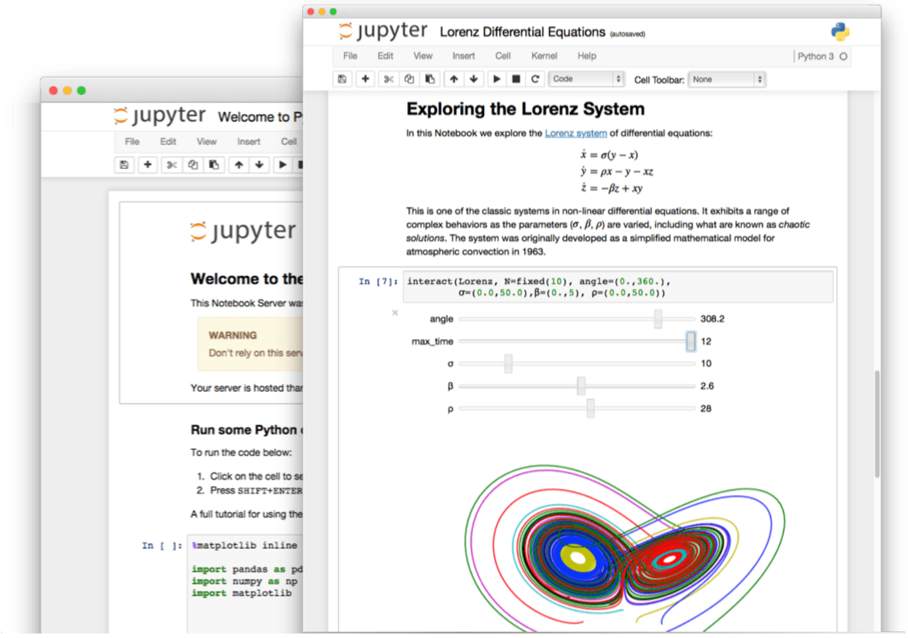

Figure 1: Example Jupyter Notebooks displaying text, code, and figures. From http://jupyter.org

1.3 Cloud computing and geosciences

In the "era of big data", the ever growing computational capacities and available data are opening many new research possibilities, but are also posing new challenges to geoscientists. Data centers are growing faster than our capacity to process and make sense of these new sources of information. A lot of resources are wasted on work duplication: the same data are downloaded and processed the same way in different institutions. Soon, it will become cheaper and faster to actually leave the data where it is (i.e. on the cloud) and bring your computing workflow to the data, rather than the other way around.

New emerging technologies are making this transition to cloud computing much easier, especially for geoscientists without formal training in computer sciences. One of these tools is Pangeo, a community platform for big data geoscience (pangeo.io). Pangeo can be seen as a community developed set of tools and recommendations about how to set-up an interactive computing environment in the cloud.

"In the cloud" means that the computer resources and the data are made available by a cloud provider (either commercial or public). Cloud computing environments are usually slower and less effective than a typical high performance cluster, but they benefit from economy of scale and allow interactive and collaborative workflows which aren't possible in private clusters. They also avoid data duplication, and give the administrators full control over the computing environment used to run a particular model.

2 Project goals

In this project, we propose to follow two main objectives:

1. **Set-up and deploy the Open Global Glacier Model in a scalable cloud environment, and make this computing environment available for everyone.**
2. **Create and run an educational platform about glaciers in this environment, allowing students of all levels to realize glacier modelling experiments and learn from them.**

With **Objective 1**, we envision an online platform where people can log-in and get access to a fully functional computing environment where they can run the OGGM model. This environment will scale according to resources demand. It will be personalized and persistent, so that a user can prepare some computations, start them, log out, then log back in and still find the computing environment he or she left earlier. The advantages for the user will be important: scalable computing resources, no installation burden, no data download, no version issues, user-friendly development environment, all in a web browser.

With **Objective 2**, we envision to use this platform not only for research but also for education and scientific communication. Our existing web page (OGGM-Edu, edu.oggm.org) is a proof of concept demonstrating how freely available cloud resources can provide the interactivity we need. In this project, we would like to develop the educational material further, with more tutorials targetting students at different levels: high-school, undergraduates, graduates.

3 Tools and methods

The central tool for setting-up the cloud environment will be Pangeo (pangeo.io). Pangeo itself relies on several services: Docker for containers, Repo2Docker for container configuration, Kubernetes for scalable deployment, JupyterHub for user management and notebooks, Python and its scientific ecosystem for the programming environment. For OGGM-Edu, we will continue to use the following tools: ReadTheDocs for the webpage, Jupyter Notebooks for the glacier experiments, Bokeh for interactive plots, GitHub for collaborative work. We cannot describe all these tools in detail here, but we are confident that they will allow us to realize our objectives within the project duration time.

For the choice of the cloud provider, we will explore several possibilities. Ideally, we would use a university provided cloud, but the LFU currently doesn't provide such a service. Therefore, we will use a commercial cloud provider (Amazon, Azure, Google...). They provide free grants for research and education, and we expect to have enough free resources to develop a proof-of-concept within the project.

4 Contribution of the project to the Digital Sciences at LFU

With this project, the LFU will support a modern and innovative approach for the future of numerical modelling and education. The educational platform will be used for or own teaching at ACINN and possibly in other departments as well (Geography, Geology). The beginner tutorials, which are oriented towards younger students, might raise their interest for studying in Innsbruck. We expect the project to yield at least one publication in "Geoscientific Model Development" or "Geoscience Communication". Importantly, the project is going to foster a collaboration between three university departments (ACINN, Digital Sciences, Computer Science) and will allow three master-level students and one PhD student to work and learn in a motivating environment. On the long term, it might stimulate the shift towards a "university cloud" at LFU, by demonstrating its usefulness for research and education. Finally, this project will demonstrate the engagement of LFU for open-science, open-knowledge and open-source development.

5 Human Resources

For the completion of this project we can rely on a strong team of LFU and international collaborators.

Fabien Maussion is an assistant professor at LFU (fabienmaussion.info). He will coordinate the project. He is employed at the University of Innsbruck since 2014 and is on a tenure track position ("Qualifizierungsvereinbarung") since Oct 2015. He has substantial experience in glaciological modelling and open source development. His contributions the domain of digital sciences include the development of high-impact open-source softwares, the creation of digital learning platforms, and teaching using innovative and interactive tools (see website and CV). The proposed research will be an important step towards the completion of his habilitation thesis.

Matthias Dusch is a PhD student at LFU, working under the supervision of Fabien Maussion and Kurt Nicolussi. He is an experienced python developer and an expert of the OGGM model. He will be involved in all aspects of the project.

Reto Stauffer is a newly appointed University Assistant in the Digital Science Center at LFU (retostauffer.org). He is an expert in statistical methods, machine learning, and web application development. He will provide guidance for the creation of the online platform and will co-supervise the master students.

Thomas Fahringer is the head of the Distributed and Parallel Systems Group at the Institute of Computer Science, LFU (dps.uibk.ac.at/ tf). He is an expert in cloud and parallel computing. He will provide guidance for the set-up of the cloud infrastructure and will co-supervise the master students.

Ryan Abernathey is an assistant professor in the Columbia University Department of Earth and Environmental Sciences, USA (rabernat.github.io). **Joe Hamman** is a Computational Hydroclimatologist working at the National Center for Atmospheric Research, USA (joehamman.com). Joe and Ryan are experts in scientific python software development and are leading the Pangeo project, a community platform for Big Data geoscience funded by the US NSF EarthCube Program. They will provide guidance and assistance to set-up the cloud infrastructure.

6 Project Duration and Requested Resources

We expect the project to run for 12 months. The requested resources sum up to **€20.510**. With this budget, we intend to hire three student helpers (master level, 8 Hours per week for 12 months, i.e. €6.837 per student). One student will be hired to set-up the cloud computing environment. Ideally, she/he will have a computer sciences background and work closely with Thomas Fahringer (Distributed and Parallel Systems Group). Two students will be hired to develop the educational web platform OGGM-Edu. They will have a background in geosciences with a solid knowledge in computer programming. Since we will use only open-source software and freely available web services, no other costs will apply.

We expect to start the project as soon as funding is available. The work on OGGM-Edu will be incremental and regular throughout the year. For the set-up of the cloud environment, we expect the system to run within 3 to 6 months, after which a period of testing will be necessary. The system will be maintained and adapted as the number of users grows.

References

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., and Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth, Nature Geoscience, p. 1, doi:10.1038/s41561-019-0300-3, 2019.

Giesen, R. H. and Oerlemans, J.: Calibration of a surface mass balance model for global-scale applications, The Cryosphere, 6, 1463–1481, doi:10.5194/tc-6-1463-2012, 2012.

Giesen, R. H. and Oerlemans, J.: Climate-model induced differences in the 21st century global and regional glacier contributions to sea-level rise, Climate Dynamics, 41, 3283–3300, doi:10.1007/s00382-013-1743-7, 2013.

Huss, M. and Farinotti, D.: Distributed ice thickness and volume of all glaciers around the globe, Journal of Geophysical Research: Earth Surface, 117, F04 010, doi:10.1029/2012JF002523, 2012.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Frontiers in Earth Science, 3, 1–22, doi:10.3389/feart.2015.00054, 2015.

Kaser, G., Cogley, J. G., Dyurgerov, M. B., Meier, M. F., and Ohmura, A.: Mass balance of glaciers and ice caps: Consensus estimates for 1961-2004, Geophysical Research Letters, 33, 1–5, doi:10.1029/2006GL027511, 2006.

Kaser, G., Grosshauser, M., and Marzeion, B.: Contribution potential of glaciers to water availability in different climate regimes, P. Natl. Acad. Sci. Usa., 107, 20 223–20 227, doi:10.1073/pnas.1008162107, 2010.

Marzeion, B., Hofer, M., Jarosch, A. H., Kaser, G., and Mölg, T.: A minimal model for reconstructing interannual mass balance variability of glaciers in the European Alps, The Cryosphere, 6, 71–84, doi:10.5194/tc-6-71-2012, 2012a.

Marzeion, B., Jarosch, a. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, The Cryosphere, 6, 1295–1322, doi:10.5194/tc-6-1295-2012, 2012b.

Marzeion, B., Cogley, J. G., Richter, K., and Parkes, D.: Attribution of global glacier mass loss to anthropogenic and natural causes, Science, doi:10.1126/science.1254702, 2014a.

Marzeion, B., Jarosch, A. H., and Gregory, J. M.: Feedbacks and mechanisms affecting the global sensitivity of glaciers to climate change, The Cryosphere, 8, 59–71, doi:10.5194/tc-8-59-2014, 2014b.

Marzeion, B., Kaser, G., Maussion, F., and Champollion, N.: Limited influence of climate change mitigation on short-term glacier mass loss, Nature Climate Change, 8, doi:10.1038/s41558-018-0093-1, 2018.

Maussion, F., Butenko, A., Eis, J., Fourteau, K., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.0, Geoscientific Model Development Discussions, pp. 1–33, doi:10.5194/gmd-2018-9, 2018.

Pfeffer, W. T., Arendt, A. a., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., Hagen, J.-O., Hock, R., Kaser, G., Kienholz, C., Miles, E. S., Moholdt, G., Mölg, N., Paul, F., Radić, V., Rastner, P., Raup, B. H., Rich, J., and Sharp, M. J.: The Randolph Glacier Inventory: a globally complete inventory of glaciers, Journal of Glaciology, 60, 537–552, doi:10.3189/2014JoG13J176, 2014.

Radić, V. and Hock, R.: Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise, Nature Geoscience, 4, 91–94, doi:10.1038/ngeo1052, 2011.

Radić, V. and Hock, R.: Glaciers in the Earth's Hydrological Cycle: Assessments of Glacier Mass and Runoff Changes on Global and Regional Scales, Surveys in Geophysics, 35, 813–837, doi:10.1007/s10712-013-9262-y, 2014.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0, Tech. rep., Global Land Ice Measurements from Space, Colorado, USA, doi:10.7265/N5-RGI-60, 2017.