The Journal of Open Source Software

auto-icon: seamless management of ICON model runs with Autosubmit

Andreas Baer [1,2¶], **Oriol Tintó Prims** [3,4], **Fabian Jakub** [3], **Miguel Castrillo** [4], **Robert Redl** [3], **René Caspart** [1], **and Martina Klose** [2]

1 Scientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany **2** Institute of Meteorology and Climate Research Troposphere Research, Karlsruhe Institute of Technology, Karlsruhe, Germany **3** Meteorological Institute, Ludwig-Maximilians-Universität München, Munich, Germany **4** Earth Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Jed Brown

Reviewers:

- @dostuffthatmatters

Submitted: 30 July 2024
Published: unpublished

Summary

Weather and climate modeling usually entails a large set of tasks, such as the retrieval of input data from repositories, the actual simulation or standard post-processing and visualization of the output data. Most of these tasks recur for each experiment and require only minor user input and no user intervention, allowing for a high degree of automation. The workflow manager Autosubmit is specifically designed to automate weather and climate modeling workflows, but it is model independent and hence needs an application interface for use with a certain model. auto-icon provides this interface, i.e. the task descriptions and configurations to run workflows with the ICON weather and climate model using Autosubmit as backend.

Statement of need

The ICON (ICOsahedral Nonhydrostatic) model (https://icon-model.org, e.g.(Zängl et al., 2014)) is a widely used high-performance modeling framework for flexible applications in weather, climate and environmental prediction. A few tools exist to conduct simulations such as BaCy, Numex, Spice (Rockel & Geyer, 2023), and the ESMtools (Gierz et al., 2021). These are all tailored to specific needs and the runs themselves are started by executing a shell script that calls a set of further scripts to perform tasks like preparation of work directories, writing of namelists, and submitting jobs to the batch system on the respective high-performance computing (HPC) system. Another common approach is the execution of a runscript, a self-contained shell script that performs these tasks for a single ICON run. If required, this script includes a loop to split the run and do reinitializations in between. This approach brings clear limitations to all aspects of the FAIR principles (GO FAIR, 2024) as there is no mechanism for storing of the workflow data after manipulation by the user. Further, there is no automation for obtaining or creating input data, post-processing or analysis. A workflow management system, capable of running such tasks and automatically documenting the workflow, can overcome these limitations.

The workflow manager Autosubmit (Manubens-Gil et al., 2016), provides a set of core functionality for workflow design and data provenance aspects to support management of workflows according to the FAIR principles (Leo et al., 2023). A clear separation of concerns between the management and execution of the workflow (Autosubmit) and the description of the workflow with the individual tasks (auto-icon) is ensured and enhances modularity and sustainability in the development process. auto-icon provides the workflow configuration and task templates to launch a full ICON model run with Autosubmit. Dependencies between tasks determine the workflow. Autosubmit then executes the individual tasks, by integrating the

provided user configuration and specific requirements of the target platform, submitting them to the batch system of the chosen HPC computing platform and monitoring their execution. The available set of tasks include building the model, preparation of input data obtained from public repositories or shared folders and preprocessing of this data for the specific run, setting up of the computational environment, and running the model. Further tasks allow to check the model output against a reference for binary identity, the visualization of core model results or archiving on a storage system available on the HPC cluster or the machine running Autosubmit. An overview of a workflow example is sketched in Figure 1, demonstrating the tasks automated by auto-icon. auto-icon provides a broad range of automation features like the addition of output required for reinitializations if required, finding of input data for the start date of the simulation and inserting input file names into the namelist. Whereas we expect these features to satisfy most use cases, the user can always overwrite them in case they require deviating settings.

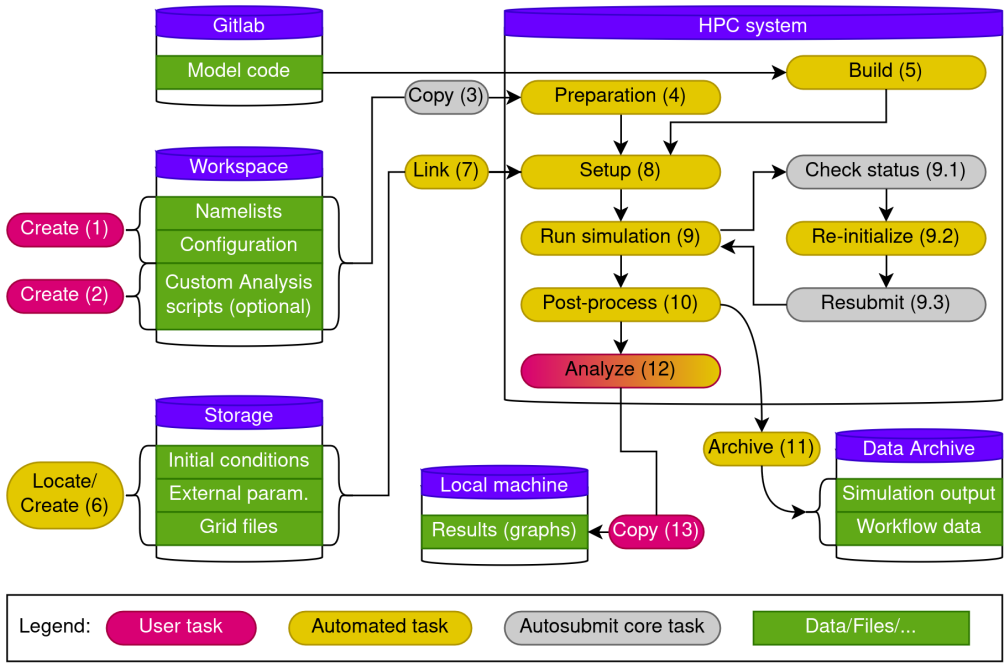

Figure 1: Basic scheme for a full auto-icon workflow including preparation and post-processing tasks. The workflow (dependencies between tasks) is created by Autosubmit. The yellow tasks are described by auto-icon and executed by Autosubmit, whereas the grey tasks are core functionality of Autosubmit.

While the experiment is running, its status is monitored by Autosubmit and can be requested by the user through the command line interface of Autosubmit and a powerful web interface (Uruchi et al., 2021). The latter provides multiple options for visualizing the experiment of interest, showing the current status of the tasks in a list, tree or graph view. Further, logs and statistics of all finished tasks can be accessed, and the configuration is also provided.

auto-icon comprises of bash and python scripts to execute the individual tasks as well as an extensive set of YAML configuration files to adapt the tasks to the variety of usage scenarios. Additionally, templates for the individual configuration to be adapted by the user are provided. The full configuration for an experiment is included in the local clone of the auto-icon repository, such that it can easily be shared and made publicly available for reproduction. Autosubmit can further provide metadata for such workflows.

Future extensions of auto-icon are planned with focus on two different aspects. The first one is to allow creation of further input data. This includes the retrieval of initial conditions

from non-standard repositories or grids and grid-based external parameters for cases where no publicly available grid data can be used, e.g. when custom nests are required. The second aspect is on the validation of simulation output. While basic validation is already possible via an external tool performing sanity checks on the output data, evaluation against observational data and different model runs is a demanded topic that can be well integrated into the workflow.

Acknowledgements

This work has received funding through the Helmholtz Association's Initiative and Networking Fund (grant agreement no. VH-NG-1533) and the Transregional Collaborative Research Center SFB/TRR 165 "Waves to Weather", funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation, Project-ID 257899354–TRR165). Parts of this work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

References

Gierz, Paul, Andrés-Martínez, M., Barbi, D., Wieters, N., Cristini, L., Streffing, J., Wahl, S., Ural, D., Kjellsson, J., Chegini, F., & Khosravi, S. (2021). *ESM-tools*. Zenodo. https://doi.org/10.5281/ZENODO.3737927

GO FAIR. (2024, July 19). *FAIR principles*. https://www.go-fair.org/fair-principles/

Leo, S., Crusoe, M. R., Rodríguez-Navas, L., Sirvent, R., Kanitz, A., De Geest, P., Wittner, R., Pireddu, L., Garijo, D., Fernández, J. M., Colonnelli, I., Gallo, M., Ohta, T., Suetake, H., Capella-Gutierrez, S., Wit, R. de, Kinoshita, B. P., & Soiland-Reyes, S. (2023). *Recording provenance of workflow runs with RO-crate*. arXiv. https://doi.org/10.48550/ARXIV.2312.07852

Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., Mula-Valls, O., & Doblas-Reyes, F. J. (2016). Seamless management of ensemble climate prediction experiments on HPC platforms. *2016 International Conference on High Performance Computing & Simulation (HPCS)*, 895–900. https://doi.org/10.1109/HPCSim.2016.7568429

Rockel, B., & Geyer, B. (2023). *SPICE (starter package for ICON-CLM experiments)*. Zenodo. https://doi.org/10.5281/ZENODO.5910888

Uruchi, W., Castrillo, M., & Beltrán, D. (2021). Autosubmit GUI: A javascript-based graphical user interface to monitor experiments workflow execution. *Journal of Open Source Software*, *6*(59), 3049. https://doi.org/10.21105/joss.03049

Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2014). The ICON (ICOsahedral non-hydrostatic) modelling framework of DWD and MPI-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 563–579. https://doi.org/10.1002/qj.2378