Potential of knowledge discovery using workflows implemented in the C3Grid

T. Engel[(1)], A.H. Fink[(1)], U. Ulbrich[(2)], T. Schartner[(2)], A. Dobler[(2)], B. Fritzsch[(3)], W. Hiller[(3)], B. Bräuer[(3)]

(1) Institute for Geophysics and Meteorology, University of Cologne (2) Institute for Meteorology, Free University of Berlin

(3) Alfred-Wegener-Institute, Helmholtz Centre for Polar and Marine Research

Introduction

With the increasing number of climate simulations, reanalyses and observations, **new infrastructures to search and analyse distributed data are necessary**. In recent years, the Grid architecture became an important technology to fulfill these demands. For the German project „**Collaborative Climate Community Data and Processing Grid**“ (**C3Grid**, **www.c3grid.de**), computer scientists and meteorologists developed a system that offers its users a webinterface to search and download climate data and use implemented analysis tools (called **workflows**) to further investigate them. In the following, two workflows of the C3Grid will be presented to show their potential to extract hidden information of extensive climate simulations.

C3Grid architecture

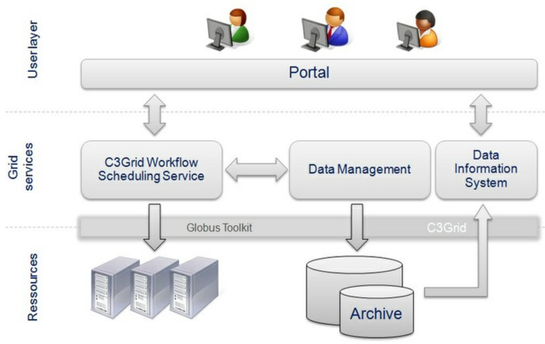

Fig. 1: Central components of the C3Grid

- User submits data or workflow request via a webportal
- Data Management Service (DMS) provides metadata of the available datasets
- Workflow Specification Language (WSL) describes the workflow
- Workflow Scheduling Service (WSS) handles the incoming requests
- Data Provider transfers requested data to Computer Provider via GridFTP
- The Compute Provider executes the workflow tasks
- Upcoming version: use of data in the Earth System Grid Federation (**ESGF**)

Workflow „Cyclone Tracking“

- Based on Murray and Simmonds, 1991, *Aust. Met. Mag.*, **39**, 155-166
- Identifying and tracking local minima in the mean sea level pressure (MSLP) field
- Adjustable thresholds:
  - Curvature of the isobars (in terms of $(\nabla^2 p_{msl})_{min}$)
  - Minimum lifetime (Δt_{min}) of a cyclone

  → **Distinction between weak subtropical heat low systems and stronger midlatitude cyclones** (e.g. Northern Atlantic)
- Output: track data including statistics (track density, average central pressure, average central curvature, cyclogenesis and cyclolysis) and built-in visualizations

Workflow „Stormtracks“

- Standard deviation of the geopotential height (σ_{GPH}) on pressure level (here: 500 hPa)
- 2.5-6 day bandpass-filtered σ_{GPH}
- Maximum σ_{GPH} → maximum synoptic activity
- Relationship to surface cyclones: stormtracks are displaced equatorwards

Example of cyclone track and stormtrack workflows

Input data:
- **MPI-ESM-LR RCP8.5** (2070-2099) and **Historical** (1970-1999)

Cyclone Tracking (Figure 2):
- Maximum cyclone activity over the North Atlantic between Newfoundland and Iceland
- Increase of cyclones over the Labrador sea, the southern Iberian peninsula and northern Africa
- Increase applies especially for weak and moderate systems ($(\nabla^2 p_{msl})_{min} < 0.8\,hPa/(deg.lat.)^2$)
- Decrease of number of tracks over the Atlantic and Central Europe
- Lifetime threshold:
  - decrease in the number of identified cyclones with increasing lifetime
  - small changes between RCP8.5 and Historical run

Stormtracks (Figure 3):
- Maximum synoptic activity over Newfoundland
- Broad area of increased synoptic activity over the Atlantic and Europe
- Decreasing activity north of ~60°N

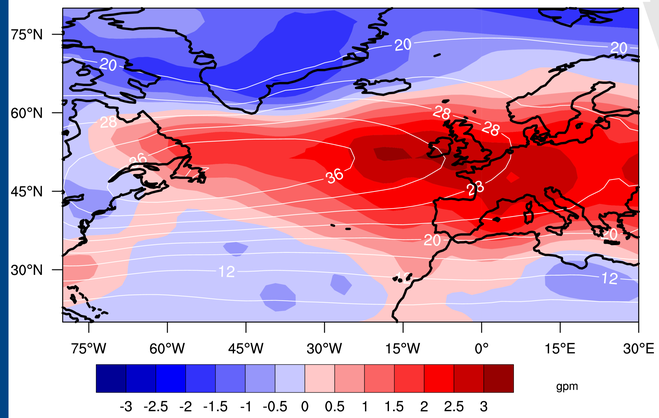

Fig. 3: As figure 2, but for 2.5-6 day bandpass-filtered standard deviation of 500 hPa geopotential height. Red (blue) shading indicates an increased (decreased) synoptic activity in the RCP8.5 scenario.

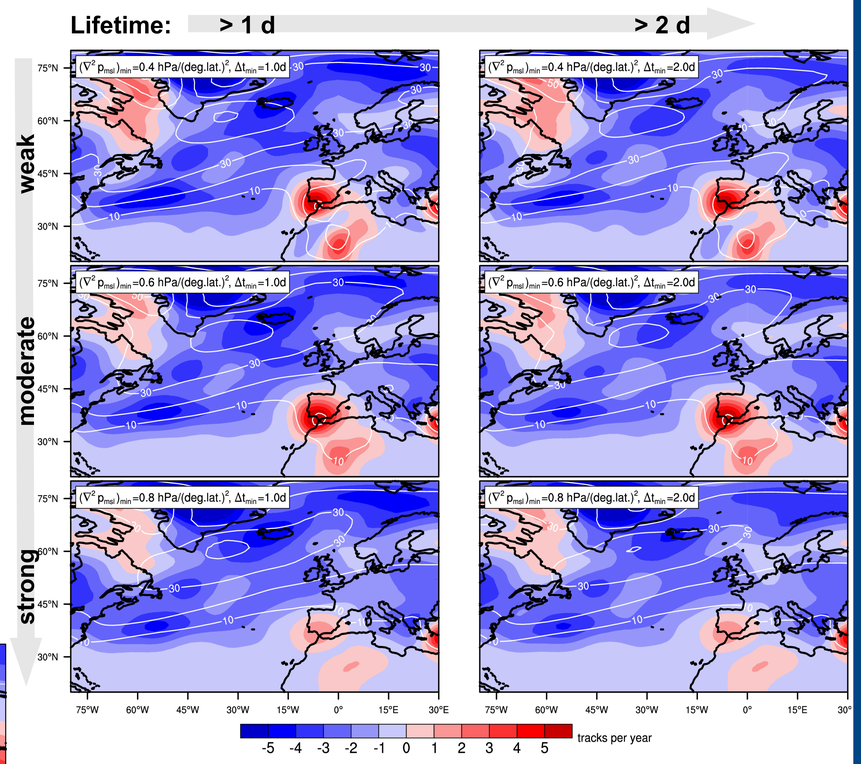

Fig. 2: Difference of cyclone tracks per year (shaded) between MPI-ESM-LR RCP8.5 (2070-2099) and Historical (1970-1999), average number of cyclones per year of RCP8.5 (contours) using different parameters of $(\nabla^2 p_{msl})_{min}$ and Δt_{min}. Red (blue) shading indicates an increased (decreased) number of cyclone tracks in the RCP8.5 scenario.

Upcoming questions:
- What causes the increased number of cyclone tracks?
- Why do the results of the cyclone tracking and stormtrack workflows differ?
- How do the results differ for other experiments (RCP4.5, RCP2.6) and other models?

Outlook on future work

The **C3Grid** will be accessible soon for public users including further workflows for the calculation of **CAPE, humidity fluxes, circulation weather types, trough identification, preparation of regional model forcing data** etc. Data of the Earth System Grid Federation (**ESGF**) and further diagnostic and utility workflows will be available. Techniques and tools to perform automated analyses of large datasets will be further developed. This will allow for batch processing of different **CMIP5** datasets using the same workflow as well as **parameter studies** varying the parameters of a single workflow. With this, the C3Grid is an efficient environment for users of the geoscience community.

Universität zu Köln

Freie Universität Berlin

AWI Alfred-Wegener-Institut für Polar- und Meeresforschung in der Helmholtz-Gemeinschaft

funded by Bundesministerium für Bildung und Forschung

Contact Poster: Thomas Engel, tengel@meteo.uni-koeln.de
Contact Project: Bernadette Fritzsch, bernadette.fritzsch@awi.de
www.c3grid.de