Helmholtz Analytics Framework
Data Analysis Methods Workshop
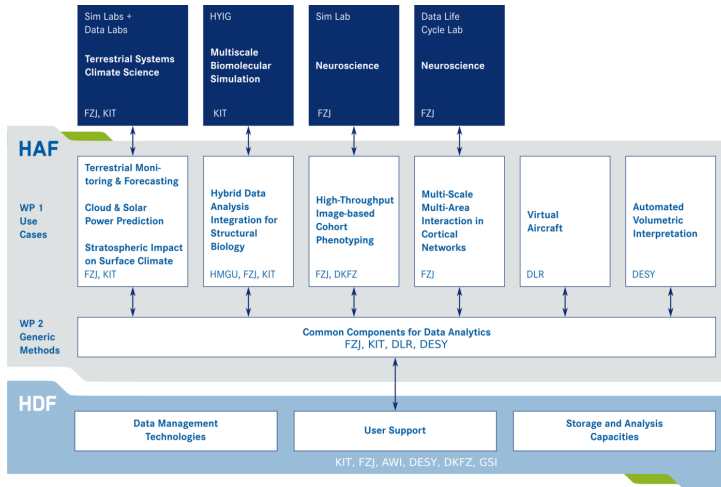
2018-03-22 | Björn Hagemeier | Juelich Supercomputing Centre

HELMHOLTZ
Analytics Framework

Helmholtz Analytics Framework

The project

- Duration: originally 3 years, October 2017 – September 2020
- Now extended until **March 2021**
- Funding: 3M€ Helmholtz Initiative and Networking Funds
  + 3M€ own contribution by participating centres
- $\Rightarrow$ 23 FTE over 3 years
- 10 FTE Generic tools
- 13 FTE Domain science

HELMHOLTZ
Analytics Framework

Project Structure

HELMHOLTZ
Analytics Framework

Project

Schedule

- We are currently in month 6
- Project extension to 42 months already granted

	month 1-6	month 7-18	month 19-30	month 31-36
WP1	identification of common methods and tools	generalizing tools and methods	use of generic methods on HDF and uptake of methods from other use cases	fully using methods on HDF
WP2		roll-out of common tools on HDF for participating use-cases	uptake of methods by further use-cases outside of this project	general availability on HDF

M1   M2   M3

HELMHOLTZ
Analytics Framework

Project
Tools

- GitLab
  - `https://gitlab.version.fz-juelich.de/haf`
  - Request JSC account at: `https://dspserv.zam.kfa-juelich.de/Dispatch/trunk/WEB/WebServices/Public/register_for_webservices.php` or
  - JSC → "JSC online, application forms" → "Apply for a Web Service Account at JSC"
- BSCW
  - `https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/2446147`
  - By invitation → Björn Hagemeier
- Mailing list
  - haf@fz-juelich.de
  - Subscribe haf-subscribe@fz-juelich.de

HELMHOLTZ
Analytics Framework

Website

www.helmholtz-analytics.de

`http://www.helmholtz-analytics.de/`

- Project information
- Use cases
- Methods
- Publications, news, events
- Change requests to Björn Hagemeier (JSC)

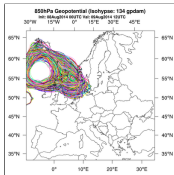**HELMHOLTZ**
Analytics Framework
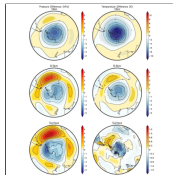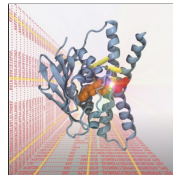
Use Cases

Overview
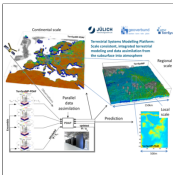
Source: IBG-3
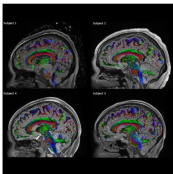Use Case 1: Terrestrial Monitoring and Forecasting

Source: IEK-8
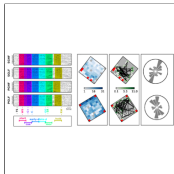Use Case 2: Cloud and Solar Power Prediction

Source: IMK-ASF
Use Case 3: Stratospheric Impact on Surface Climate

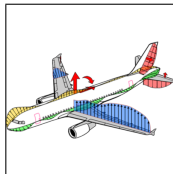Source: SCC
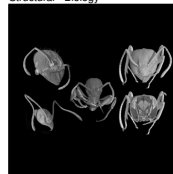Use Case 4: Hybrid Data Analysis and Integration for Structural Biology

Source: INM-1
Use Case 5: High-Throughput Image-Based Cohort Phenotyping

Source: INM-6
Use Case 6: Multi-Scale Multi-Area Interaction in Cortical Networks

Source: Airbus
Use Case 7: Virtual Aircraft

Source: Nova Project
Use Case 8: Automated Volumetric Interpretation

HELMHOLTZ
Analytics Framework

Generic Methods

- General Machine Learning methods
  - Clustering, classification, hyper parameter optimization
- Made available on HDF and HPC resources
- Derive methods from use cases
- Cross-fertilization among use cases
- Benchmark method implementations on various underlying libraries
- Provide advice to use cases
  - performance tuning
  - alternative methods
- Produce library/framework to support use cases, HGF, and others

HELMHOLTZ
Analytics Framework

Agenda

Today

Time	Topic
09:00 - 09:30	Welcome
09:30 - 11:00	Clustering, K-Means, DBSCAN
11:00 - 11:30	Break: Discussion
11:30 - 12:30	Clustering: Self-organizing Maps
12:30 - 13:30	Break: Lunch
13:30 - 15:00	Classification: Logistic Regression
15:00 - 15:30	Break: Discussion
15:30 - 17:00	Classification: Neural Networks

HELMHOLTZ
Analytics Framework

Agenda
Tomorrow

Time	Topic
09:00 - 10:30	Data Assimilation: Ensemble Kalman Filter
10:30 - 11:00	Break: Discussion
11:00 - 12:30	Data Assimilation: 4D Var
12:30 - 13:30	Break: Lunch
13:30 - 15:00	Sequence Mining: SPADE, FP-Growth
15:00 - 15:30	Break: Discussion
15:30 - 17:00	Hyperparameter Optimization: Grid-Search, Bayesian Optimization
17:00 - 17:30	Outlook

HELMHOLTZ
Analytics Framework

Introduce Yourself

Get to know each other

Name	Björn Hagemeier
Institute	Forschungszentrum Jülich
	Juelich Supercomputing Centre (JSC)
Background	Computer Science
Role in the project	Project management
	Generic methods

HELMHOLTZ
Analytics Framework

Questions

HELMHOLTZ
Analytics Framework