ISDA-Online

January 14, 2022 / 07 – 09 UTC

“Data Assimilation Infrastructure - Software, Frameworks, HPC”

Organizers: *Javier Amezcua (U. Reading, UK)*
Ting-Chi Wu (RIKEN, Japan)
Tobias Necker (U. Vienna, Austria)

Program: Six talks with each 18 min. (15 min. talk + 3 min. questions)

07:00 – 07:04	**Welcome**
07:04 – 07:22	**The Data Assimilation Research Testbed in 2022 and Beyond: Not Your Parents' DART** Moha Gharamti, Helen Kershaw, and DAReS Team at NCAR
07:22 – 07:40	**A global 3.5km-mesh 1024-member ensemble data assimilation benchmark - Our challenges on the supercomputer Fugaku** Hisashi Yashiro, Koji Terasaki, Takemasa Miyoshi, and Hirofumi Tomita
07:40 – 07:58	**1000-member 30-second-update real-time precipitation forecast experiment using the supercomputer Fugaku** Arata Amemiya, Takumi Honda, Shigenori Otsuka, Yasumitsu Maejima, James Taylor, Hirofumi Tomita, Seiya Nishizawa, Kenta Sueki, Tsuyoshi Yamaura, Yutaka Ishikawa, Shinsuke Satoh, Tomoo Ushio, Kana Koike, Erika Hoshi, and Takemasa Miyoshi
07:58 – 08:02	**Time buffer for questions**
08:02 – 08:20	**Melissa-DA: An Elastic and Fault Tolerant Large-Scale Online Data Assimilation Framework** Sebastian Friedemann, Kai Keller, Yen-Sen Lu, Bruno Raffin, and Leonardo Bautista-Gomez
08:20 – 08:38	**Developing a common, flexible and efficient framework for weakly coupled ensemble data assimilation based on C-Coupler2** Chao Sun, Li Liu, Ruizhe Li, Xinzhu Yu, Hao Yu, Biao Zhao, Guansuo Wang, Juanjuan Liu, Juanjuan Liu, Fangli Qiao, and Bin Wang
08:38 – 08:56	**PDAF - community software for ensemble-based data assimilation** Lars Nerger
08:56 – 09:00	**Closing: Information on upcoming events**

Please note:

- **The times in UTC are approximate. In case of technical problems, we might have to change the order of the presentations.**
- When you login to the session before 07:00 UTC, and everything is quiet, this is most likely because we muted the microphones.
- **Time Zones: 07 – 09 UTC**
  Europe: 07 – 09 am GMT (London) | 08 – 10 am CET (Berlin)
  Asia/Australia: 03 – 05 pm CST (Shanghai) | 04 – 06 pm JST (Tokyo) | 06 – 08 pm AEDT (Sydney)
  Americas: 11pm – 01 am PST (San Fran.) | 00 – 02 am MST (Denver) | 02 – 04 am EST (New York)

The Data Assimilation Research Testbed in 2022 and Beyond:
Not Your Parents' DART

Moha Gharamti, Helen Kershaw and DAReS Team
NCAR Data Assimilation Research Section

The mission of the Data Assimilation Research Section (DAReS) at NCAR is to accelerate progress in actionable Earth System Science at NCAR, UCAR universities and in the broader science community by providing state-of-the-art ensemble DA capabilities including DA for NCAR community modeling systems. For nearly 20 years, the DAReS team has done this using the Data Assimilation Research Testbed (DART), an open-source community facility that provides software tools for data assimilation research, development and education. While the goals have remained the same, DART has evolved continuously to address novel science applications using state-of-the-art software on cutting edge computing platforms.

The current processes for adding new models and observations to DART will be reviewed along with examples of several recently developed applications with novel scientific impact. A review of the science results will demonstrate some of the many available DART diagnostic tools. DART uses specialized ensemble filter algorithms that allow a single source to run efficiently on platforms ranging from laptops to supercomputers. This framework also allows straightforward, efficient implementation of novel nonlinear and non-Gaussian ensemble filters in addition to hybrid ensemble-variational variants. These algorithms and their parallel implementation will be discussed along with a preview of current development exploring the use of GPUs in preparation for exascale computing technology. Pointers to the comprehensive DART documentation and a variety of educational/tutorial information will be provided.

A global 3.5km-mesh 1024-member ensemble data assimilation benchmark - Our challenges on the supercomputer Fugaku

Hisashi YASHIRO[1], Koji Terasaki[2], Takemasa Miyoshi[2], and Hirofumi Tomita[2]

[1] National Institute for Environmental Studies (NIES), Japan
[2] RIKEN Center for Computational Science, Japan

The ground-breaking, massive ensemble data assimilation (DA) benchmark has been conducted on the Japanese new flagship supercomputer "Fugaku" using Nonhydrostatic ICosahedral Atmospheric Model (NICAM) and Local Ensemble Transform Kalman Filter (LETKF). Fugaku has a many-core general-purpose CPU (A64FX) based on the Arm instruction set. Also, Fugaku has a large memory bandwidth comparable to GPU, but the memory capacity per node is small. Therefore, it was necessary to reduce the memory footprint used by each computational process. The data transfer, including file I/O, was also a bottleneck in the execution time of the whole DA cycle. We applied a data-centric design approach to the NICAM-LETKF system such as; 1) parallel file input/output for the exchange of the data between the simulation and the DA, 2) usage of fast distributed local storage, 3) reduction of global communication, 4) usage of less precision floating-point variables and operation. These software optimizations/developments were conducted as a systems-applications co-design to achieve 100 times higher performance than the K computer. We achieved good computational scalability both in the atmospheric simulation and DA. The development of a numerical software library specialized for single-node single-precision calculation has succeeded in significantly reducing the time for eigenvalue calculation, which increases as the number of ensemble members increases. Finally, we realized a global 3.5 km mesh, 1024-member ensemble data assimilation with 131,705 nodes (6,291,456 cores) of Fugaku. In this experiment, a 1.3 PB of data was transferred from the simulation part to the DA part. And we showed that about four hours are required to complete one DA cycle.

1000-member 30-second-update real-time precipitation forecast experiment using the supercomputer Fugaku

Arata Amemiya[1,2], Takumi Honda[1], Shigenori Otsuka[1,2], Yasumitsu Maejima[1], James Taylor[1], Hirofumi Tomita[1,2], Seiya Nishizawa[1], Kenta Sueki[1,2], Tsuyoshi Yamaura1, Yutaka Ishikawa[3], Shinsuke Satoh[4], Tomoo Ushio[5], Kana Koike[6], Erika Hoshi[6], Takemasa Miyoshi[1,2,7]

[1] RIKEN center for computational science
[2] RIKEN Cluster for Pioneering Research
[3] National Institute of Informatics
[4] National Institute of Information and Communication Technology
[5] Osaka University
[6] MTI Ltd.
[7] Interdisciplinary Theoretical and Mathematical Sciences Program

The supercomputer Fugaku is Japan's flagship supercomputer deployed in 2020 as the successor of the previous K computer. Fugaku won the first place in four rankings of supercomputers (TOP500, HPCG, HPL-AI, Graph500) in November 2021 for the fourth consecutive term. Once Fugaku started its full operations in March 2021, we conducted a real-time data assimilation and numerical weather prediction experiment with unprecedentedly large size to demonstrate Fugaku's capability. In the selected periods in the summer of 2021 during the Tokyo Olympic and Paralympic games, we ran the regional data assimilation and numerical weather prediction system SCALE-LETKF assimilating radar reflectivity and Doppler velocity of the multi-parameter phased array weather radar (MP-PAWR) every 30 seconds. The LETKF calculation was performed with 1000 members and extended 30 minutes forecasts were performed with 10 members for each initial time. We used 10,810 nodes in total, which corresponds to about 8 % of the entire Fugaku system. During most of the experimental periods, we successfully provided deterministic and probabilistic 30 minutes precipitation forecast refreshed every 30 seconds in Tokyo area and disseminated them on our website and smartphone application. The presentation will include the technical challenges and achievements in our experiments in performing real-time data assimilation with 1000 members using Fugaku.

Melissa-DA: An Elastic and Fault Tolerant Large-Scale Online Data Assimilation Framework

Sebastian Friedemann[1], Kai Keller[2], Yen-Sen Lu[3], Bruno Raffin[1], and Leonardo Bautista-Gomez[2]

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
[2] Barcelona Supercomputing Center, 08034 Barcelona, Spain
[3] Forschungzentrum Juelich, 52428 Juelich, Germany

Simulating chaotic systems (e.g., hydrological, climate or weather systems) requires both, observations and numerical-model states. As such model's resolution and simulation volume increase, capturing the fundamental nonlinear dynamics, including the system's full error covariance, becomes impractical. Ensemble-based data assimilation (DA) is a prevailed, however costly tool propagating the system together with the flow-dependent covariance. We present Melissa-DA, an elastic, online, fault-tolerant and modular framework for large-scale ensemble-based DA. The framework allows elastic addition and removal of allocated computing resources for state propagation at runtime. Data exchange between the components relies on the network layer and not on MPI, making the framework tolerant towards failures of individual components. Multiple members are propagated by each model instance allowing to minimize resource requirements. Melissa-DA provides an interface to connect climate models and assimilation techniques, includes a built-in interface to PDAF and its parallel-DA filter methods, and provides its own particle filter implementation. Experiments confirm the excellent scalability, running EnKF on up to 16,240 cores, propagating 16,384 members for a regional hydrological critical zone assimilation relying on the ParFlow model on a domain with about 4 M grid cells.

Developing a common, flexible and efficient framework for weakly coupled ensemble data assimilation based on C-Coupler2

Chao Sun (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China)
Li Liu (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China)
Ruizhe Li (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China)
Xinzhu Yu (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China)
Hao Yu (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China)
Biao Zhao (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China; First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China; Key Lab of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao, China)
Guansuo Wang (First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China)
Juanjuan Liu (State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China)
Fangli Qiao (First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China; Key Lab of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao, China)
Bin Wang (Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China; State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China)

Data assimilation (DA) provides better initial states of model runs by combining observational information and models. Ensemble-based DA methods that depend on the ensemble run of a model have been widely used. In response to the development of seamless prediction based on coupled models or even earth system models, coupled DA is now in the mainstream of DA development. In this paper, we focus on the technical challenges in developing a coupled ensemble DA system, which have not been satisfactorily addressed to date. We first propose a new Data Assimilation Framework based on C-Coupler2 (DAFCC) for weakly coupled ensemble DA, which enables users to conveniently integrate a DA method into a model as a procedure that can be directly called by the model. DAFCC automatically and efficiently handles data exchanges between the model ensemble members and the DA method, and enables the DA method to utilize more processor cores in parallel execution. Based on DAFCC, we then develop a sample weakly coupled ensemble DA system by combining the ensemble DA system GSI/EnKF and the coupled model FIO-AOW. This sample DA system and our evaluations demonstrate the effectiveness of DAFCC in both developing a weakly coupled ensemble DA system and accelerating the DA system.

PDAF - community software for ensemble-based data assimilation

Lars Nerger[1]

[1] Alfred Wegener Institute, Bremerhaven, Germany

PDAF, the Parallel Data Assimilation Framework (http://pdaf.awi.de), is an open-source framework for ensemble-based data assimilation (DA). PDAF is designed to be particularly easy to use and a DA system can first be built quickly and then extended to complex cases, while PDAF ensures the computational efficiency. PDAF provides the option to either use separate programs to compute the ensemble of forecasts and the DA analysis steps or to augment a model code for ensemble data assimilation so that one obtains a data-assimilative model. For this online-coupled DA, the ensemble-component of PDAF provides functionality by inserting 3 function calls into the upper-level part of a model code. These additions allow that data transfers are performed in memory and using the MPI parallelization standard, which leads to excellent parallel scalability and performance.

To apply the ensemble DA, PDAF provides DA methods, in particular ensemble Kalman filters and particle filters. Additional functionality for DA is provided in form of tools for diagnostics, ensemble generation, and for generating synthetic observations for OSSEs or twin experiments. PDAF is widely used for research purposes and teaching, but also for operational DA. The recently developed Observation Module Infrastructure (OMI) provides functionality for observation handling and a structured implementation approach to easily manage a large set of different observation types. In addition, OMI allows to generate a community collection of observation operators.

The most recent release planned for December 2021 also includes 3D variational and ensemble-variational DA solvers, additional toy models, and a revised MPI structure. Recent developments further include support for strongly-coupled DA across components of Earth system models, model bindings for NEMO, SCHISM, and the climate model AWI-CM. This presentation discusses PDAF's features and recent infrastructure developments of this community DA software.