Open Source Science for ESO Mission Processing Study

Identify a system architecture that meets the ESO mission processing objectives, supports open science, enables system efficiencies, and promotes earth-system science.

Workshop #1
October 19-20, 2021

MC Mission Data Processing System Perspective

David Wiese
Caltech/Jet Propulsion Laboratory

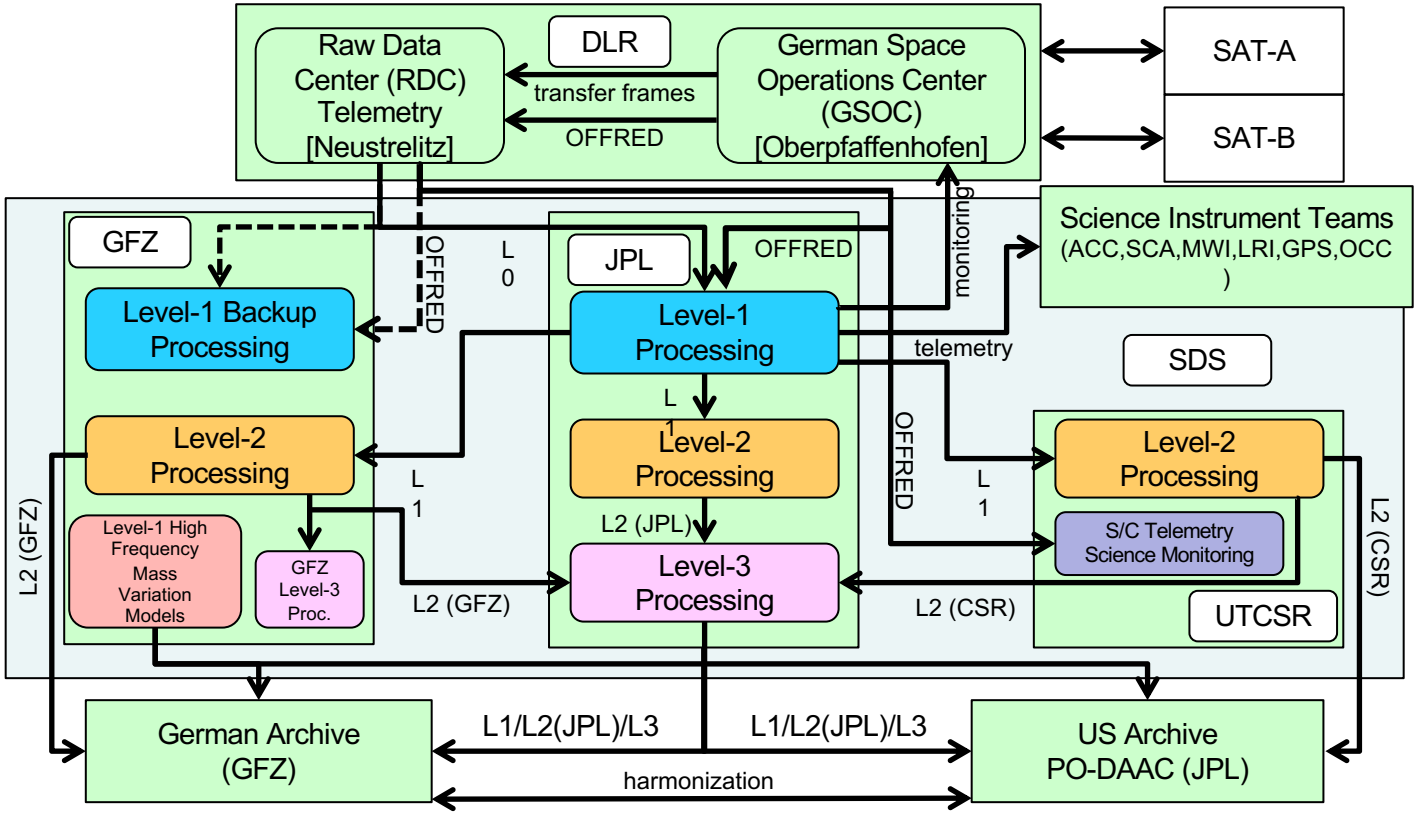
Data Processing Systems Architecture

MC inherits a mature data processing systems architecture that has successfully delivered 20 years of mass change observations.

Potential changes:
-International Partner unknown

Note: GSFC is additional L2 Processing Center for product validation and ancillary product generation for GRACE-FO (also for MC)

GRACE-FO Architecture Depicted Below



Component and Infrastructure View

Dive one step deeper than the system architecture view (previous diagram) to show:

- Algorithms:
  - Important to have unique algorithms founded on the same fundamental principles between the different L2 processing centers. This diversity is a strength and a form of cal/val.
- Computing and Storage
  - This could be common between the L2 processing centers, but in the past, each center has been responsible for their own computing and storage environment. Logistically making this common would likely be challenging with unclear benefits
- Infrastructure
  - Heritage: processing centers rely heavily on local compute nodes. Some more recent movement towards shared supercomputer systems (i.e. TACC, Pleiades)
  - Heavy computing costs have made local computing solutions more affordable in the past.
- External organizations
  - As depicted on the previous slide

Deep Dives of Components of SDS Processing Workflows are in Backup

Current Implementation Plan

Inherited software capabilities

- 20-year legacy SDS capabilities for L1 and L2 processing. Relies on software developed over the last 50 years, independently owned by each processing center. For GRACE-FO an Algorithm Theoretical Basis Document (ATBD) was released in lieu of source code which is protected under Caltech IP.

Project Schedule

- Pre-launch: Numerical simulations to verify requirements are met
- Post-launch: First data release will have high level of maturity.
    - Example: First GRACE-FO data release quality was on par with GRACE data quality after 15 years of maturation

Supporting Earth System Science

- Does your MDPS use/share any data, algorithms, etc from other ESO projects to support Earth System Science?
  - No. Primary aspect of integration with other ESO projects is with L3 gridded products; data assimilation systems
- What are your barriers to enabling collaboration to support Earth System Science within your ESO MDPS? (e.g., firewalls, access, schedule, developments costs, etc)
  - It is unclear that there are advantages to having this collaboration in the MPDS, rather than forming the collaboration outside the MDPS, based on L3/L4 data products
- What are the opportunities for improved support of Earth System Science?
  - Providing robust estimates of uncertainty can aid in data assimilation systems.
    - This is not currently neglected – attention is given to the topic, but it is a challenge
  - Harmonizing MC observations with other remote sensing observations of mass change related phenomena:
    - Earth surface deformation (NISAR/SDC, Altimetry observations (ocean/ice), GNSS surface displacements)
    - Comparisons of datasets and combinations of datasets can yield insights into Earth System

Supporting Open Science

SMD defines open science as a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding.

- What does this definition of Open Science mean in the context of data processing systems?
  - We see the primary entry point of maximum return on investment as making L3 and L4 data products widely available and accessible. 77% of PODAAC downloads of GRACE-FO data are L3/L4 data products!
- What do you feel are the most beneficial opportunities for improvement in the MDPS to support Open Science?
  - Increasing user access to instant creation of on-demand mass change timeseries with robust uncertainty information provided.
- What are the barriers to supporting open science?
  - Proprietary code in all processing centers for L1 and L2 processing. However, we do not believe there is a big return on investment with making these parts of the MDPS open. No requests from users for this code.
- What components (Data system, PGEs, algorithms, data) of your system will be developed in the open (open source from the outset)?
  - Heritage system largely relies on proprietary software with open algorithm specification for replication.

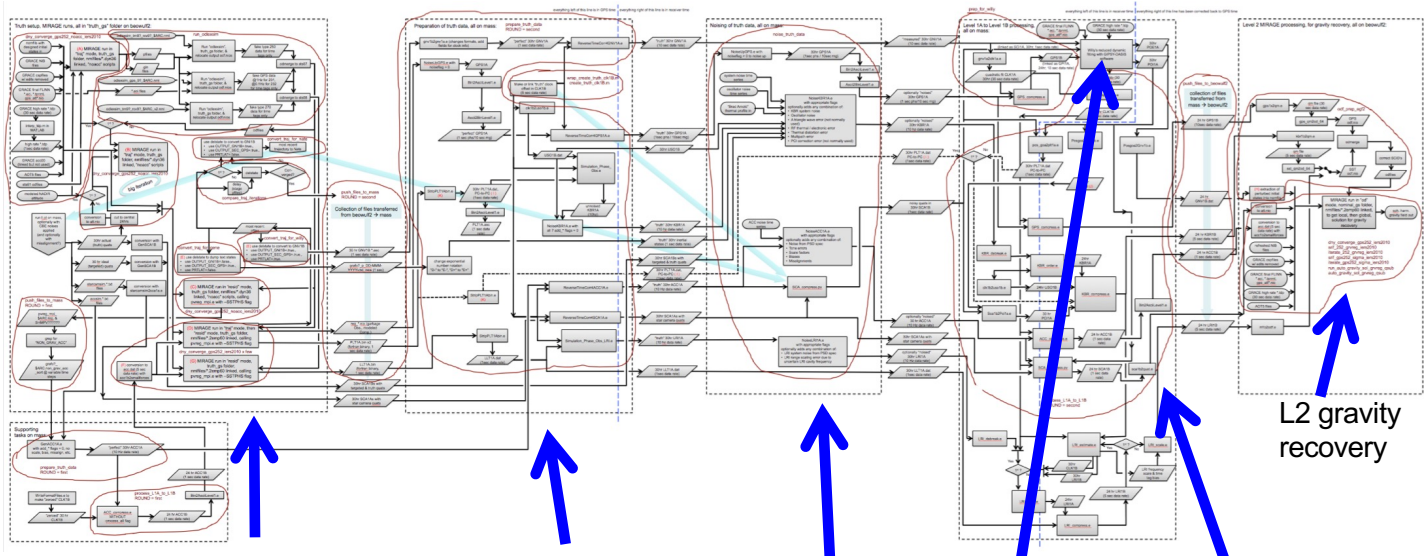
Other

- What are your pain points in support of this mission?
  - Major point of pain: maintenance and consistent reprocessing of 20+ year climate data record can strain resources; requires significant computing resources and experienced personnel with specific skills
- What does system efficiency mean in the context of an MDPS? (cost, data storage, processing time, etc.)
  - Cost-effective computing. Data processing requires significant computation time.
  - Automation to the extent possible to reduce need for human intervention
- Is there anything else you'd like to share that you feel would be helpful in our study?
  - Reiteration: Mature MDPS exists that has been finely tuned over the last 20 years of processing GRACE and GRACE-FO data. MC will leverage and build upon these systems.

BACKUP

- Following Slides show flow diagrams of processing chains for GRACE-FO.

Pre-Launch Verification of Requirements



L2 truth models generate mutually consistent truth orbits, non-gravitational acceleration and attitude time histories

Make L1A data products, apply realistic clock offset

Degrade L1A data products with CBE noise models

Perform L1 POD & clock offset recovery

L1A to L1B processing: do clock correction, form DOWR, compression

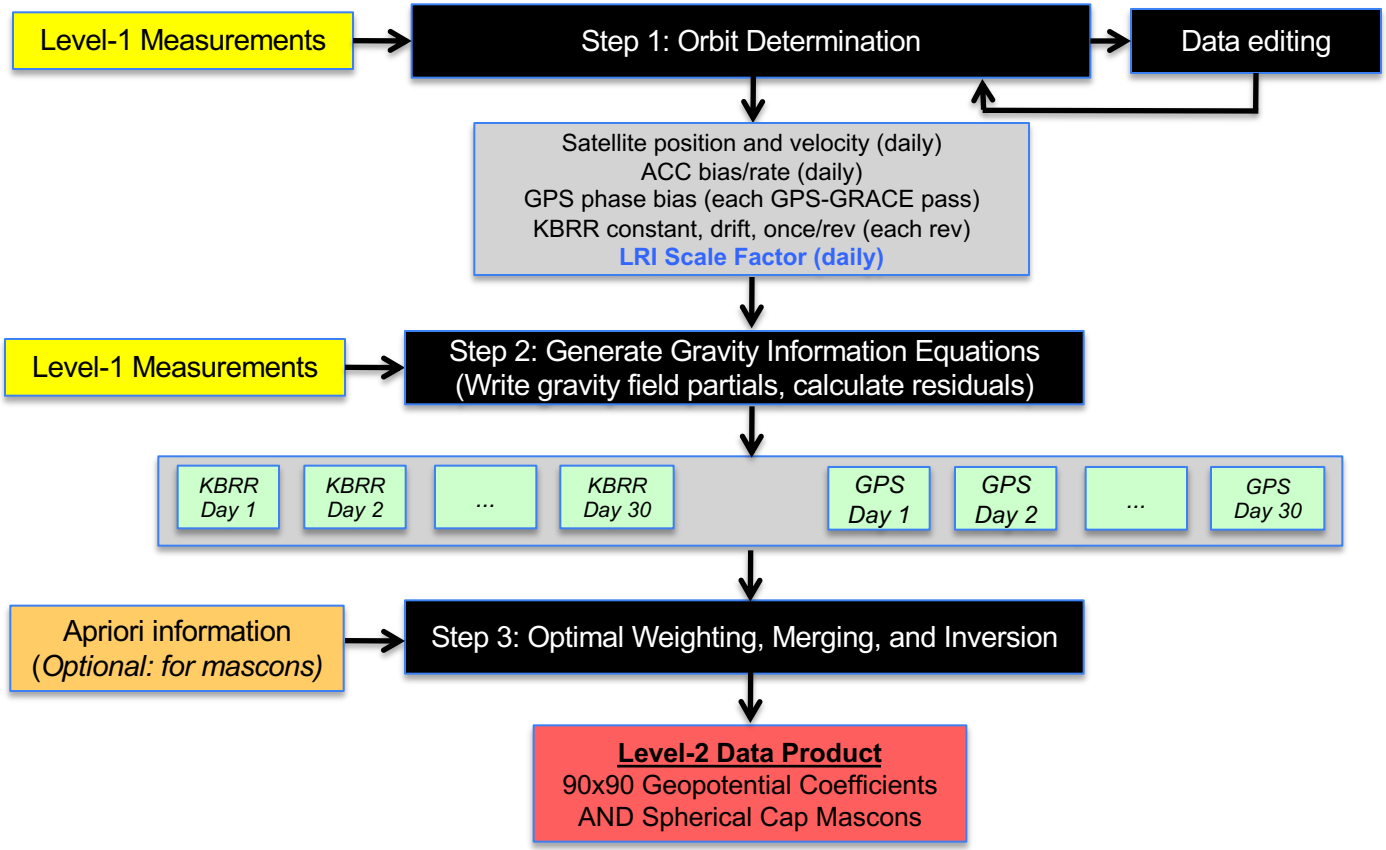
L2 gravity recovery

Grand Simulation achieved all objectives

Grand Sim objectives:

1) Provide full system test of all SDS processing software
2) V&V tool to verify GRACE-FO gravity error requirements

GRACE-FO JPL Level-2 Overview



GRACE-FO: Level-3 Design