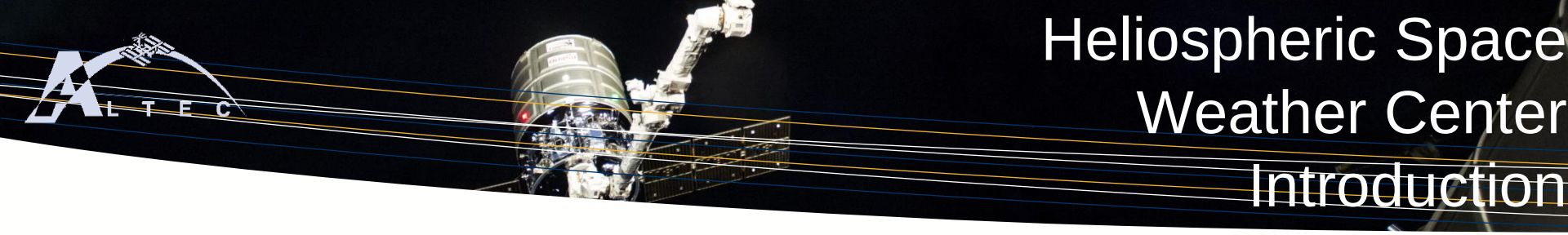
Near-Real time data management and processing system to develop and validate space weather services

***A.F. Mulone*[1]**, *M. Casti*[1], *R. Susino*[2], *R. Messineo*[1], *E. Antonucci*[2], *G. Chiesura1*, *D. Telloni*[2], *R. De March*[1], *E. Magli*[3], *A. Bemporad*[2], *G. Nicolini*[2], *S. Fineschi*[2], *F. Solitro*[1], *M. Martino*[1]

[1]ALTEC S.p.A., Corso Marche 79, 10146 Torino, Italy
[2]INAF-OATo, Via Osservatorio 20, 10025 Pino Torinese, Torino, Italy
[3]Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

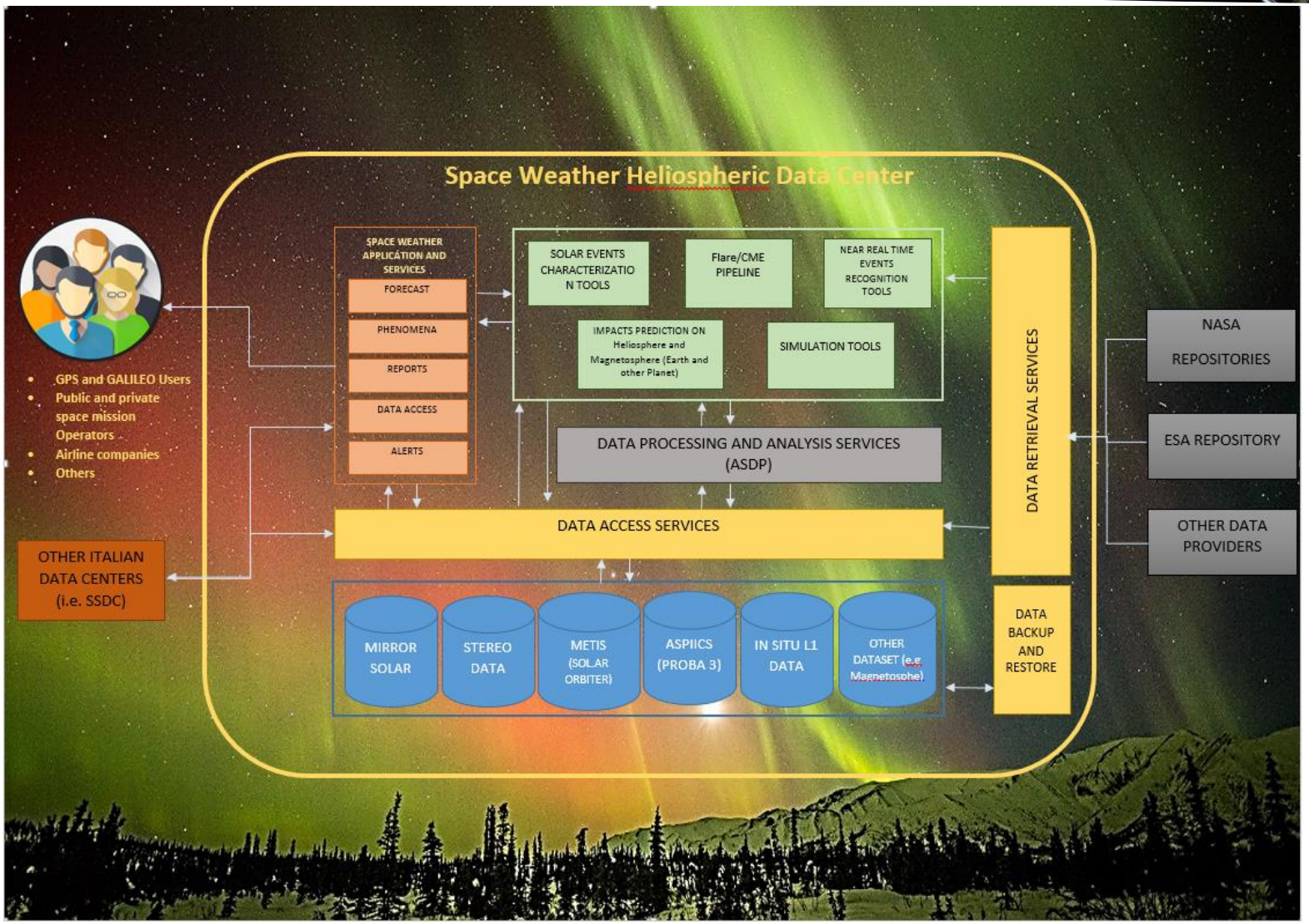
20/02/2019



Heliospheric Space Weather Center
Introduction

- Near-real time data management and processing system, designed and implemented to provide space weather services
- Joint effort between ALTEC S.p.A. and INAF-OATo
- Main goal is reducing the time between the space weather services
  - Tens of integrated data sources from past and current missions
  - Easy data integration through enriched standard data model
  - Big data technologies used to overcome data management and processing challenges
  - Integrated auxiliary tools and dataset for performing product validation.
- It is part of Heliospheric Space Weather Center for early forecast of impacts of solar disturbances on the heliosphere and the earth magnetosphere

Heliospheric Space Weather Center

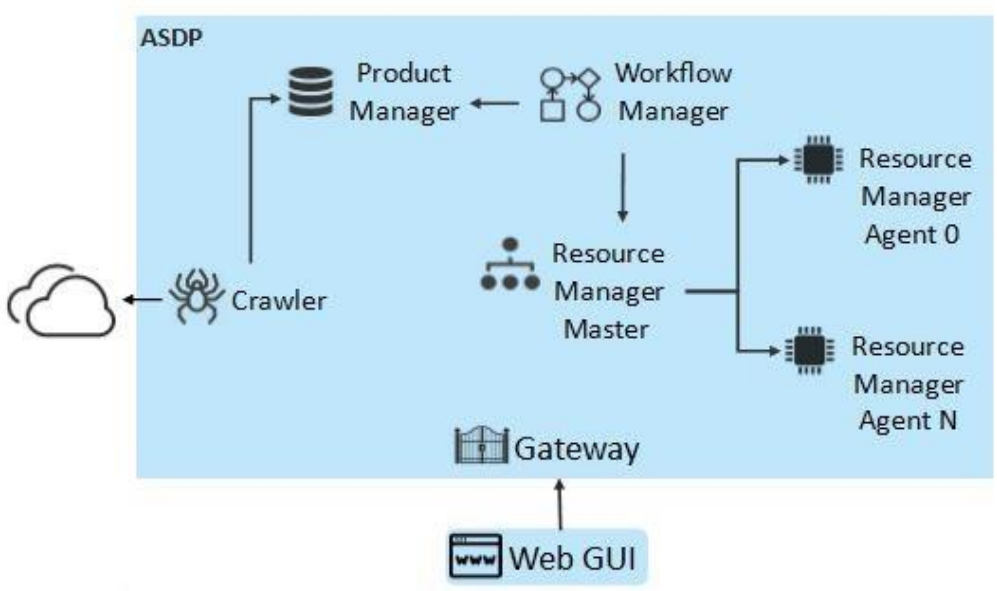


Heliospheric Space Weather Center Infrastructure

- Heliospheric Space Weather Center is deployed on ALTEC environment
- ALTEC has a private cloud on top of the Dev&Test Infrastructure.
  - The ALTEC cloud has several environment that can be used to tune software architectures
- The system provides a set of RESTful API for products, events and pipelines

Data management system consists of:

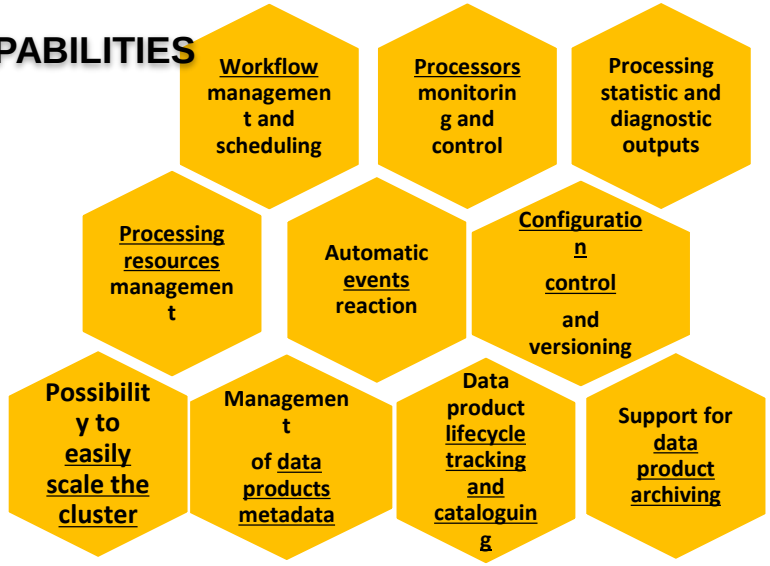
- Product Manager
  - Hands all different data stores and offers a transparent data access service
    - Data stores
    - Metadata data store
    - Processing data stores
- Crawler
- Workflow Manager
- Resource Manager
  - Master
  - Agent



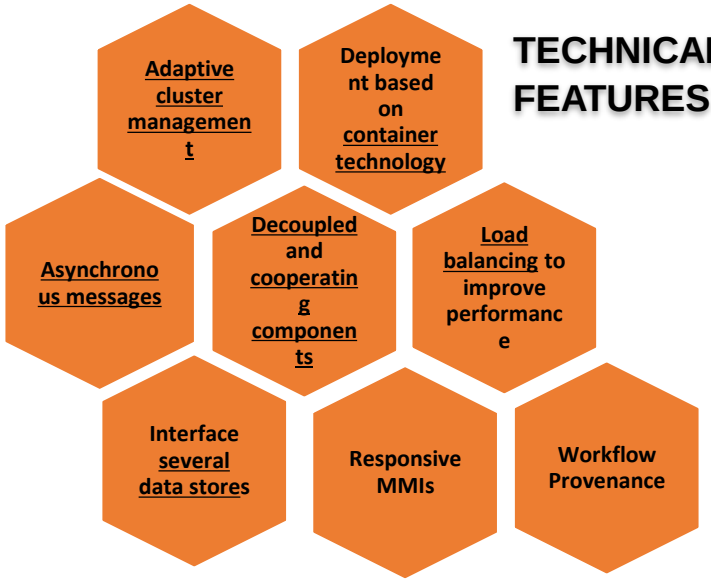
Altec Space Data Processing ASDP

- ASDP is a **distributed data processing framework** dedicated to the on-ground handling and transformation of any aircraft and spacecraft data.
- Its modular architecture gives it the flexibility to be easily adapted to several other domains where commonalities with ground centers could be found, even outside of aerospace domains.
- ASDP enables both **automatic and batch processing** of large dataset and the **automatization of complex pipelines**.

CAPABILITIES

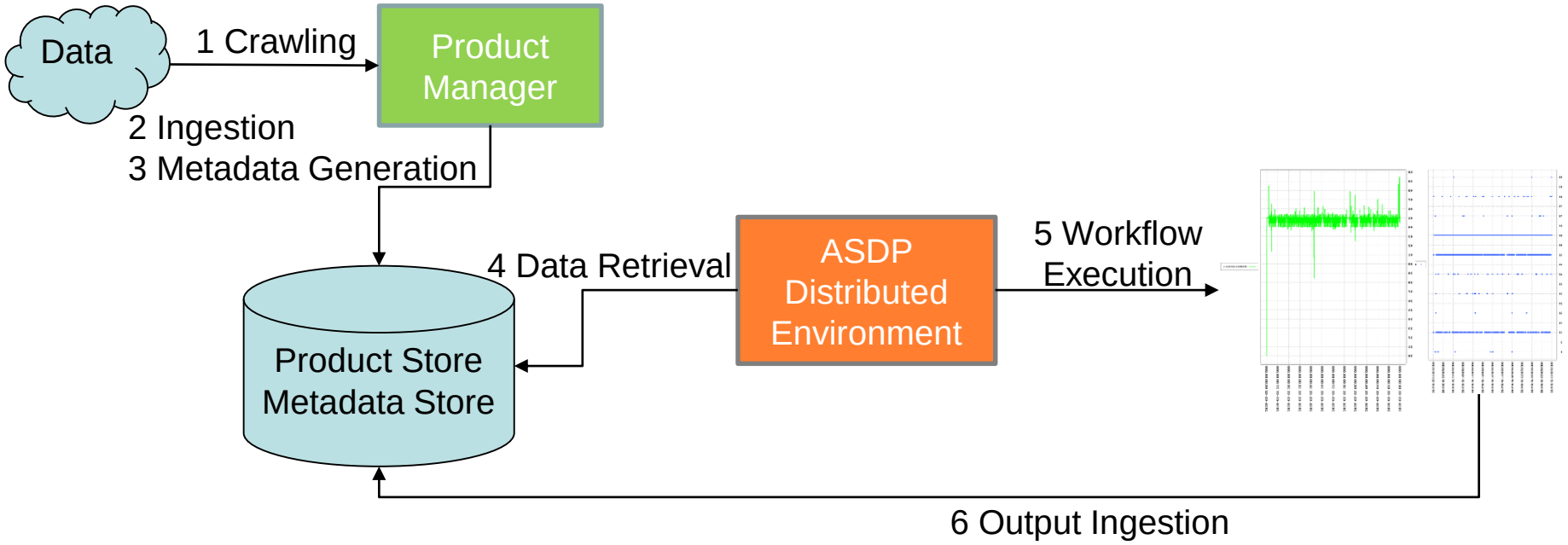


TECHNICAL FEATURES



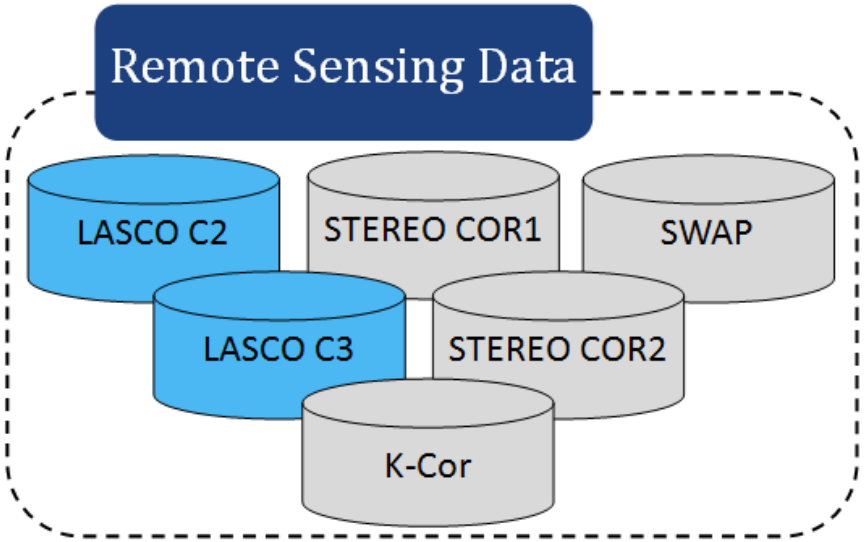
Near Real Time Data Flow

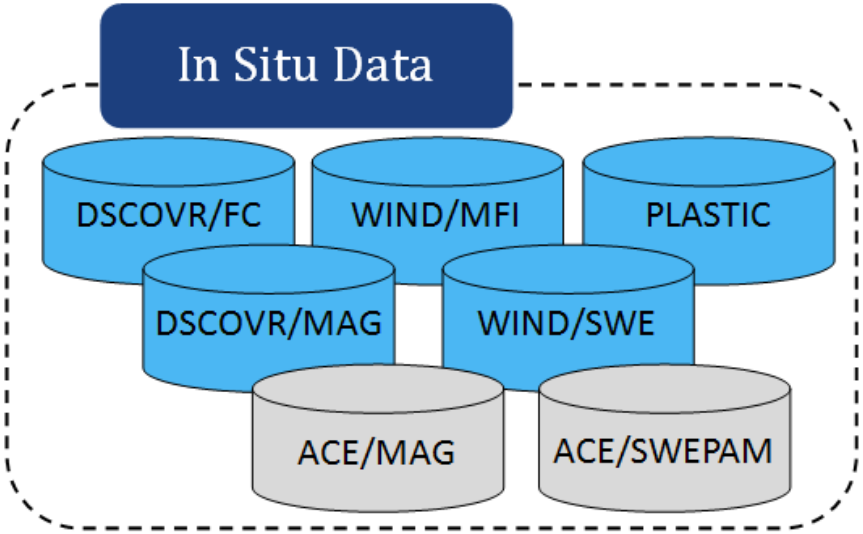
- Crawling is the capability to retrieve data from different sources
- ASDP is able to continuously check for new data and ingest new content when available.
- Metadata is generated during the product ingestion and is fundamental for data retrieval



Available Solar and Heliospheric Databases

- The Heliospheric Space Weather Center has access to several solar databases of both remote sensing and in situ data.





Data Model And Metadata

Each data received and produced in HSWC has metadata compliant with SPASE model

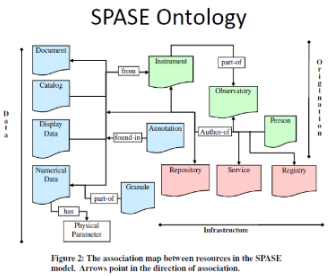
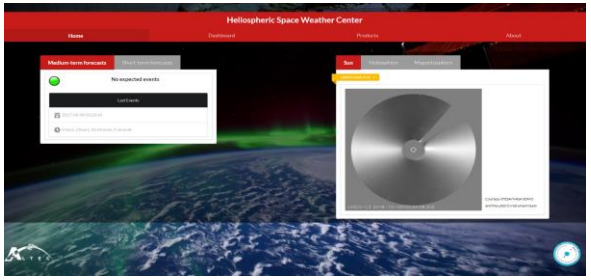


Figure 2: The association map between resources in the SPASE model. Arrows point in the direction of association.

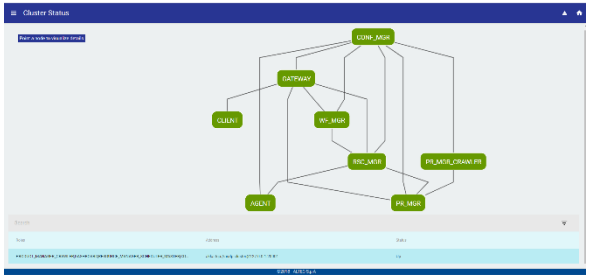
- SPASE has been extended for describing the processing of produced output
- To describe better the processing of generated output ESPAS was integrated in SPASE data model
- ESPAS represents better the processing while SPASE express better the structure
- Reference
  - http://www.spase-group.org/
  - https://www.espas-fp7.eu/

Data Access and Visualization

- Heliospheric Space Weather Center has a dedicatd interface for scientists and an operator interface



- Scientists interface
  - Homepage provides
    - Latest detected CME events
    - Latest data related to the SUN: Heliosphere and Earth Magnetosphere
  - Products page provides
    - Download of the products ingested in HDC
    - Download of the pipelines output products



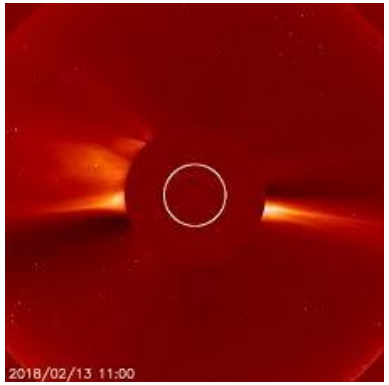
- Operator interface allow to check
  - Pipelines status
  - System interaction

- Three pipelines run in HDC
  - Remote Sensing pipeline
  - In-Situ pipeline
  - Deep Learning Pipeline
- There are integrated thanks to flexibility of ASDP
  - Each pipeline exploit different technologies and processes different data products

Data Processing Scientific Pipelines

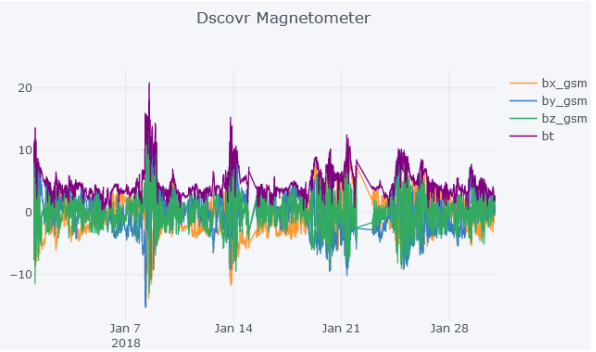
Remote Sensign Pipeline

- Exploit SolarSoft system and IDL
- Processes latest SOHO LASCO C2 and C3 fits files available
- Steps
  - Image calibration
  - Calibrated image analysis
  - Event detection
  - Retrieving of CME physical parameters
  - Propagation time at L1 and probability of impact on Earth calculus



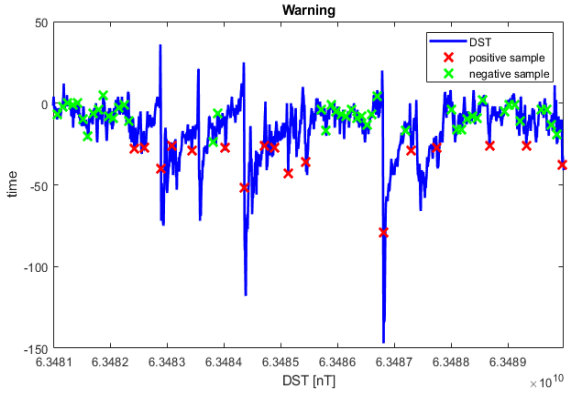
In-Situ Pipeline

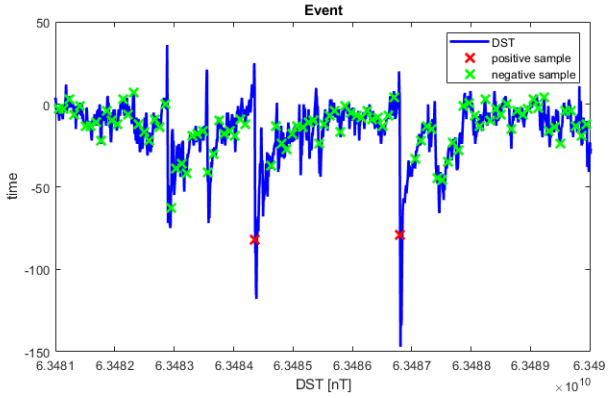
- Ingestion in real time of DSCOVR data
- Pipeline is fed with the latest 28 days of data
- Data processed in order to forecast geoeffective events



Data Processing
Deep learning pipeline

- Predict the DST geomagnetic index value that will be measured within few hours
- Use the data to derive models, and make predictions based on the models
  - Statistical models
  - Neural networks
- RNN coded in python.
- Exploit TensorFlow library
- Created a specific docker container with required libraries
- Processes real time data stored in PostgreSQL
- Trained using WIND data
- Processes the in situ data acquired by the DSCOVR payloads





Validation Resources

- Tests have been performed in order to integrate algorithms and validate the scientific results
- ASDP framework allow the creation and the execution of automatic test
  - Test script is an XML file
  - Output of the tests is an XML file
- Validation of scientific results requests
  - Development of suitable tools
  - Selection of auxiliary data sets related to near real time measurements
  - Historical data collected within CME catalogues available online
    - DST Index
    - SOHO - SoHo Long-term Archive
    - LASCO CME Catalogue (manual catalogue, out of date but with more information)
    - CACTUS Catalogue (automatic catalogue, available in real time but with less information)

Future Prospects

- Capacity to manage different data products
- Capacity to manage different data stores
- Integration of different framework and libraries
  - To improve services and algorithms
- Intregration of other data models (e.g. DACS)
- Integration of new versions of Remote Sensing, In-Situ and deep learning pipelines
- Developemnt of the cross-check between the pipelines
- Development andn integration of deep learning pipeline on image products
- Improvement of backup techniques

www.altecspace.it
angelo.mulone@altecspace.it