DAPCOM
Data services

Company presentation and possible contributions to GaiaNIR

Jordi Portell

on behalf of Francesc Julbe (CEO)

Company overview

- Funded in 2013 as a technological **spin-off** company of the **University of Barcelona** (UB) and the **Technical University of Catalonia** (UPC)
- Main motivation: Commercial exploitation of the **FAPEC data compression** patent
- Founders with strong background on **Gaia** → **software and data processing** services
- Incubated at the ESA BIC Barcelona (Castelldefels), 2014-2016
- Moved to Vic (Barcelona) in 2020
- Small company: now 8 people (2 PhD)
- Close cooperation with IEEC/ICCUB

Main products and services

- Main divisions:
  - Efficient data compression
    → **Product**/Technology (FAPEC)
  - Massive data handling and processing
    → **Service** (software engineering)

FAPEC
All-in-one
Data compression solution

- Data compression (FAPEC):
  - **Fast**, lightweight and performant
  - **Being used in ~100 Spire satellites**
  - Also used in ESA's **OPS-SAT**
  - Also being used in **maritime data**
  - Gaia DR2, EDR3 and DR3 bulk data in .csv.fapec
  - Also **WinFAPEC** (GUI) and **fapyc** (Python)

- Data visualization (FARSHY):
  - Fast viewer for hyperspectral imagery and maritime watercolumn files

FAPEC
FARSHY
Raw, Soundings and Hyperspectral

From Space to the Oceans

☐ Sponsored article

Efficient data compression and visualization software for multibeam echosounders

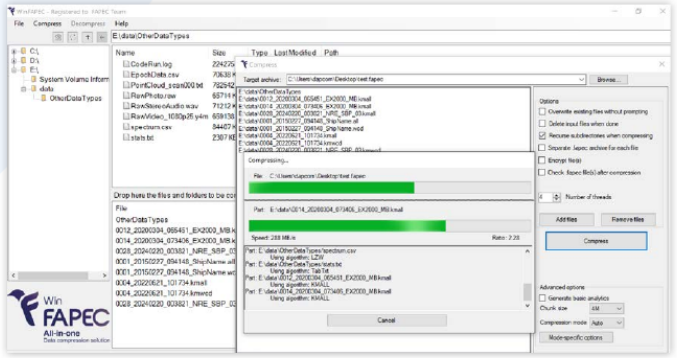
Surfing the big data wave

Water column data acquired by multibeam echosounders (MBES) imposes large requirements on disk storage and data transfer, so it is typically logged only during specific times, a practice that poses the risk of missing interesting targets. Furthermore, huge data volumes from both bathymetry and water column data can lead to considerable burdens for the operators during long surveys. MBES data is often compressed using standard solutions such as Zip or 7-Zip, but these can be computationally heavy for a relatively modest size reduction. To overcome this, we developed FAPEC, a high-performance data compression software, now supporting MBES data. We also present FARSHY, a fast visualization and analysis tool to streamline quick checks on the heavy water column files.

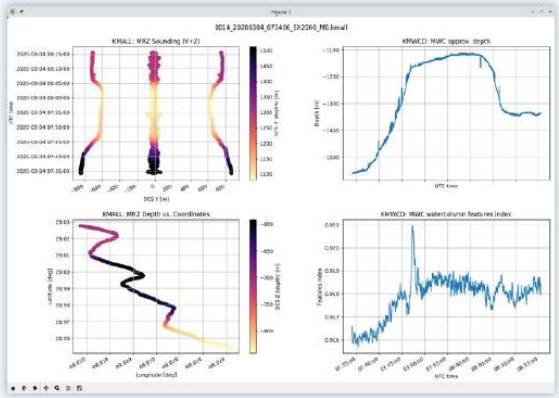
Space technology for marine echosounders

DAPCOM's FAPEC data compression software was originally designed for satellites such as ESA's Gaia, the billion-star surveyor, where onboard computing, storage and downlink capabilities are extremely limited. Later, FAPEC was enhanced with improved performance and additional algorithms to better adapt to a wide variety of file formats and data characteristics. In collaboration with Kongsberg Discovery and the Marine Geosciences Research Group of the University of Barcelona (UB), FAPEC was adapted to accommodate the *.all*, *.wcd*, *.kmall* and *.kmwcd* (KMall) formats from Kongsberg's EM MBES systems, and more recently it has been integrated into Kongsberg's Seafloor Information System (SIS) to provide automated file compression once the logging files have closed. FAPEC is being further extended to other formats and vendors.

FAPEC runs on Windows (including a graphical user interface, WinFAPEC), as well as on macOS and Linux, and it supports ARM processors. Its C, Python (*fapyc* package) and Java API allows for integration in third-party software. FAPEC rapidly examines the files to be compressed, determining the best algorithm and configuration for each of them. It supports tabulated text files (such as CSV or point clouds), multidimensional time series and multispectral images, to name a few. Therefore, rather than a *universal* data compressor, FAPEC is adaptive and versatile, allowing a much more efficient use of resources.



▲ *Figure 1: Screenshot of WinFAPEC while compressing several files on a standard laptop.*



▲ *Figure 2: Plots obtained with Python from on-the-fly basic data analytics provided by FAPEC, from soundings (left: beam width, depth and coordinates) and water column data (right: depth and features index).*

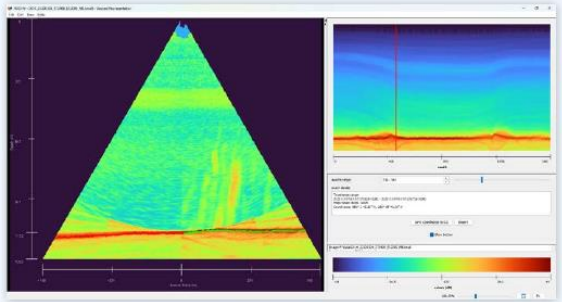
On MBES datasets kindly provided by Kongsberg and Fugro (who have started using FAPEC on their vessels), FAPEC demonstrated superior performance: it achieved better compression than *7-Zip*, while running 50 times faster and using 30 times less memory. Depending on the echosounder and scenario, FAPEC further reduced the file sizes (compared to 7-Zip) up to 10% for water column data, and up to 23% for combined bathymetry and water column data.

Beyond data compression

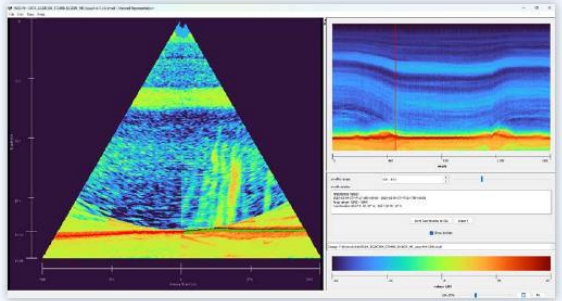
By default, FAPEC runs in lossless mode, meaning that the original files can be exactly recovered. However, for .wcd and KMall files, it also provides several lossy compression options, meaning that the quality of the data is slightly degraded to achieve a better compression. For KMall bathymetry (soundings), it allows for an instrumentally lossless operation, just removing the measurement noise. The seabed image samples can be quantized at a level indicated by the user, and can even be mostly removed if not needed. A similar approach is provided for older (.wcd) water column files.

For KMall water column data, besides sample quantization, FAPEC also provides a *smart lossy* mode, which examines the sample values and removes those below a given percentile. This makes it possible to keep most of the features in the water column (including sub-bottom data) while vastly improving the compression ratio. In the specific example shown in the FARSHY screenshots, the combined bathymetry and water column KMall file is 933MB, which is reduced to 410MB in lossless mode, and just 154MB with these lossy options. When adequately adjusted, water column files can become even smaller than bathymetry files while retaining most of the relevant information.

FAPEC achieves these results by knowing the data format and examining the values. It can provide *basic data analytics* on the fly, namely small CSV-like text files with a digest of the file contents. For example, it generates a *water column features index*, which aims to indicate sudden changes in the scene such as those created by gas seeps, fish shoals or



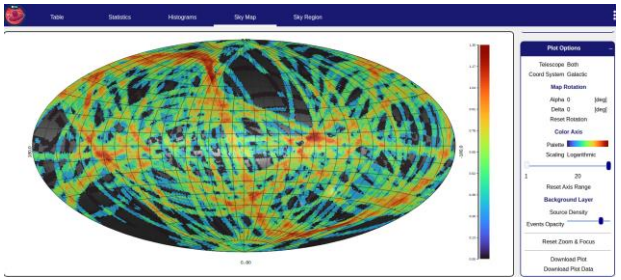
▲ *Figure 3: Screenshot of FARSHY showing gas seeps in the water column (left) and the along-track view (top right). (Data courtesy: Fugro)*

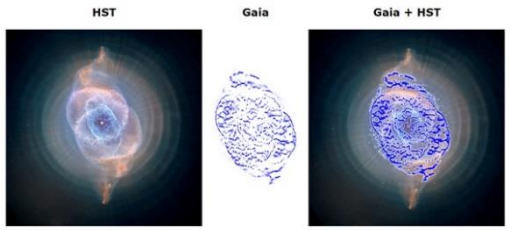


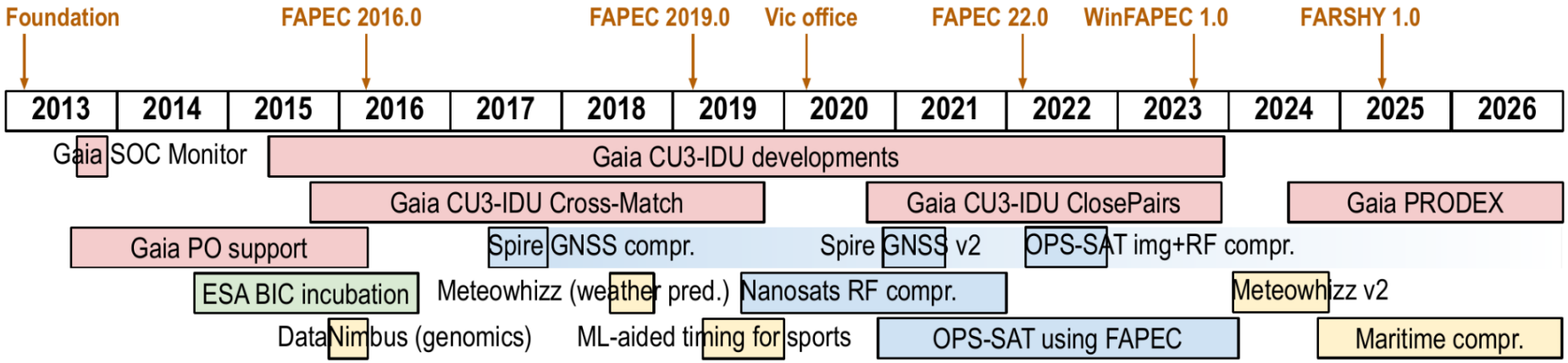
▲ *Figure 4: Same water column file as in Fig. 3 after lossy compression by FAPEC, reaching a compression ratio of six while showing an even clearer view of the gas seeps.*

Main products and services

- Some projects:
  - Touchscreen and front-end for the Gaia SOC (ESA)
  - **Gaia/DPAC IDU developments (IEEC)**
  - **Gaia/DPAC XM algorithm (ESA)**
  - **Gaia/DPAC ClosePairs resolution (ESA)**
  - **Gaia/DPAC IDU, DPCB and CU9 (ESA/PRODEX)**
  - Image & RF compression onboard OPS-SAT (ESA)
  - Weather forecast & monitoring pipeline (IAC, Meteowhizz)
  - Optical timing of runners (MACSHA)







Know-how and expertise

- **Software development**
  - ECSS standards (quality, documentation, testing) and good practices
  - Java, C/C++, C#, Python, shell scripting, GUIs
- **Storage technologies**
  - Databases
  - Efficient file formats (e.g., HDF5, Parquet)
  - On-the-fly fast compression/decompression
- **Massive data processing**
  - Tailored pipeline developments
  - Deployment and operation in HPC
  - BigData technologies: Hadoop, Spark
  - Visualization tools and techniques
- **Cloud computing** (mainly Google)
  - VMs, DataProc clusters, BigQuery, VertexAI, Docker, OpenStack
- **Algorithm development**
  - E.g., clustering algorithms (XM), attitude refinement, decorrelators/compression

What can we do for GaiaNIR

- **Software development**
  - ECSS standards (**quality, documentation, testing**) and **good practices**
  - Java, C/C++, C#, **Python**, **shell scripting**, GUIs
- **Storage technologies**
  - **Databases**
  - Efficient file formats (e.g., HDF5, **Parquet**)
  - On-the-fly fast **compression**/decompression
- **Massive data processing**
  - Tailored **pipeline developments**
  - **Deployment and operation in HPC**
  - **BigData** technologies: Hadoop, Spark
  - **Visualization** tools and techniques
- **Cloud computing** (mainly Google)
  - **VMs**, DataProc clusters, BigQuery, VertexAI, **Docker**, **OpenStack**
- **Algorithm development**
  - E.g., **clustering algorithms (XM)**, **attitude refinement**, decorrelators/**compression**

DAPCOM
Data services

Thanks!

Jordi Portell (jordi.portell@dapcom.es)

on behalf of Francesc Julbe (CEO)

www.dapcom.es / contact_us@dapcom.es