DESP forerunner: harmonised earth data provision

Alessandro Amici <a.amici@bopen.eu>

B-Open - https://bopen.eu

Company presentation

B-Open at a glance

- Technology partner for monitoring the health of the planet
    - → Python scientific and geospatial ecosystems, Open Source
    - → Earth observation, weather & climate, innovative algorithms

- Team of 20+ software engineers
    - → Ph.D. and masters in Physics, Mathematics, Engineering

- Yearly revenues around €2-3M

- 20 years experience, ISO 9001 certified, based in Rome

Main customers and projects

- ECMWF
  - → Copernicus: Climate Data Store
- ESA
  - → DestinE: Harmonised Earth Data Provision forerunner
  - → Copernicus: Transformation framework
- EUMETSAT
  - → Data Tailor Web Service
- Open source
  - → Xarray and Zarr contributors (sponsored by Chan Zuckerberg Initiative)
  - → B-Open projects: Xarray-Sentinel, Sarsen (sponsored by Microsoft)

HEDP data access forerunner

State of the art data access and transformation

Climate data store (CDS) of the Copernicus climate change service

- Quality controlled climate datasets: ERA5, SEAS, CMIP6, etc

- World scale: 250k users, 10k active users / day, 150Tb download / day

- Storage based on ECMWF MARS archive and global fields in GRIB format

- Advanced features: data selection and transformation, fair queuing

- Pain points:

  → slow access to time-series

  → complex request language

  → advanced processing means queuing
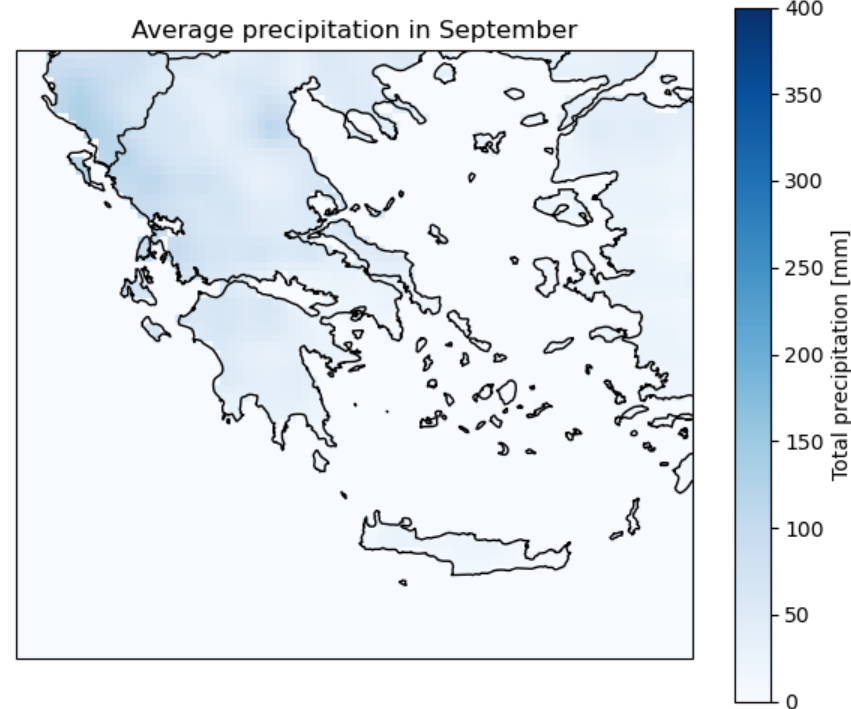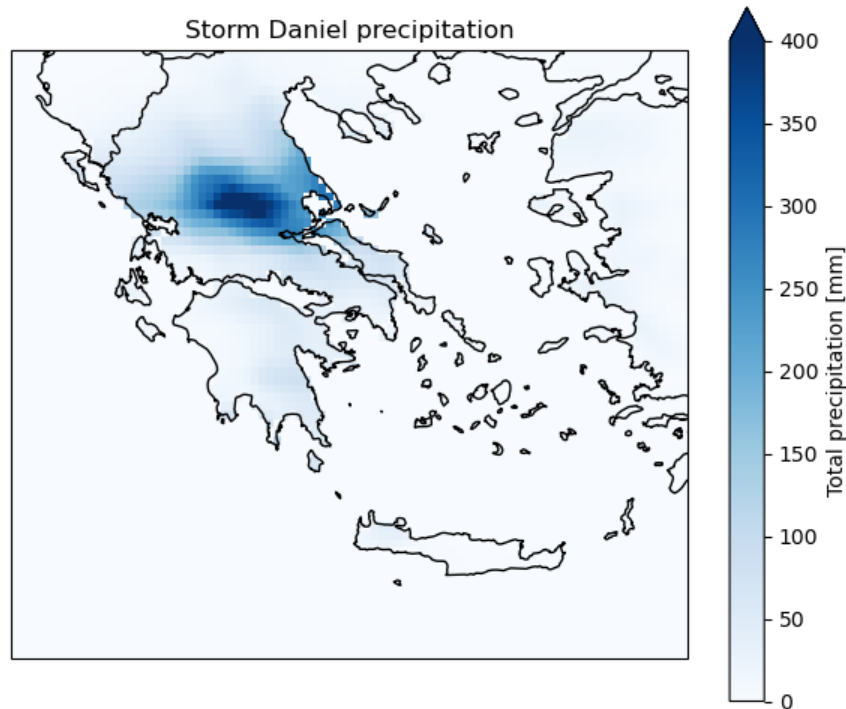
Use case: Greece floods from CDS

Map precipitation in Greece 5 and 6 September 2023 (ERA5 Land)

Storm Daniel precipitation
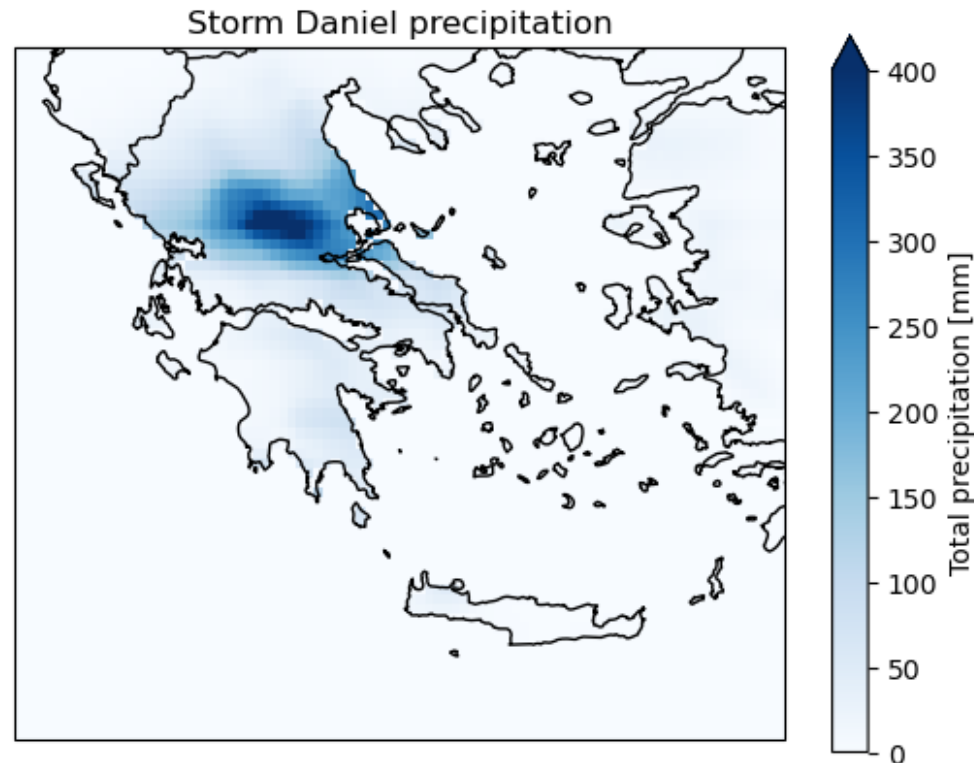
2 sec – supports interactive use
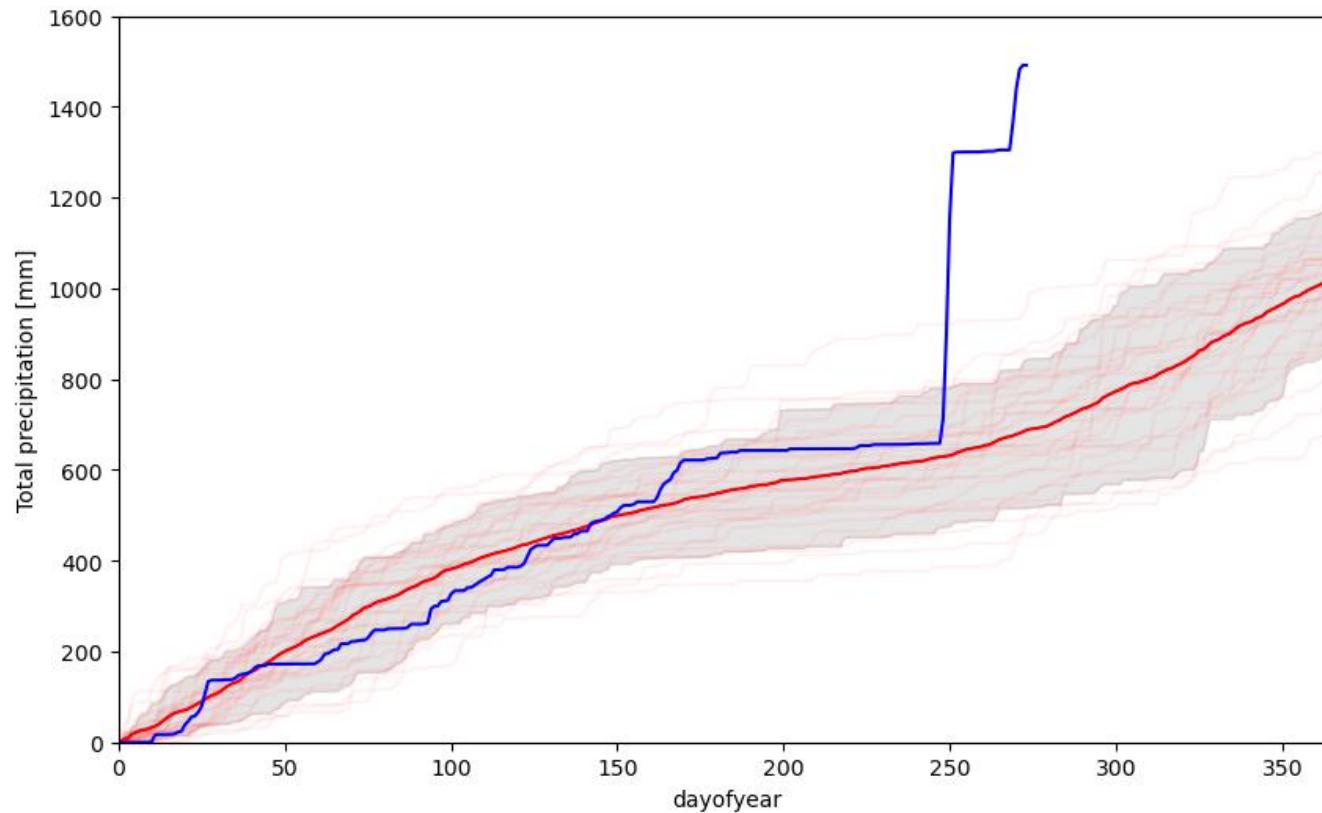
Use case: Greece floods from CDS

Compare precipitation in Greece to 1990-2019 September average (ERA5 Land)

30 min – daily reporting workflows

Use case: Greece floods from CDS

Compare accumulated precipitation over Thessaly to 1990-2019 (ERA5 Land)

7 hours – batch analysis

Challenges of weather and climate data access

- Domain needs:
  - → Petabyte scale (climate), terabyte scale with frequent updates (weather)
  - → Large user community with wildly disparate access patterns

- Technical challenges:
  - → No de-facto standard protocol for rich data access and transformation
  - → No one-size-fits-all storage granularity

- Current solutions
  - → Complex: custom data access protocols
  - → Slow for some use cases, typically time-series

DESP advanced data access for users

Goal: A optimised solution to support Platform local analysis of the data

Users needs for modern data access:

- use cloud native data formats: STAC, Zarr, Cloud optimised GeoTIFF, TileDB

- support cloud native processing: Dask, Xarray, Pandas

- simplify the user experience

- use optimal data granularity for users – support interactive time-series analysis

B

DESP advanced data access for users

Roadmap:

- Forerunner service with supporting catalogue ☑

- Integration as a core DESP service 🚧

- Add DestinE data to catalogue 🚧

- Operation in June 2024

```python
import xarray as xr

ds = xr.open_dataset(
    "s3://ecmwf-era5-land/reanalysis-era5-land.zarr", engine="zarr", chunks={}
)

location_year = ds.tp.sel(**location).sel(time=year).isel(step=23).groupby("time.time")[datetime.time()].compute(
location_reference = ds.tp.sel(**location).sel(time=reference_period).isel(step=23).groupby("time.time")[datetime

display.compare(location_year, location_reference, ylim=[0, 1600])
```
Last executed at 2023-11-14 10:19:41 in 34.75s
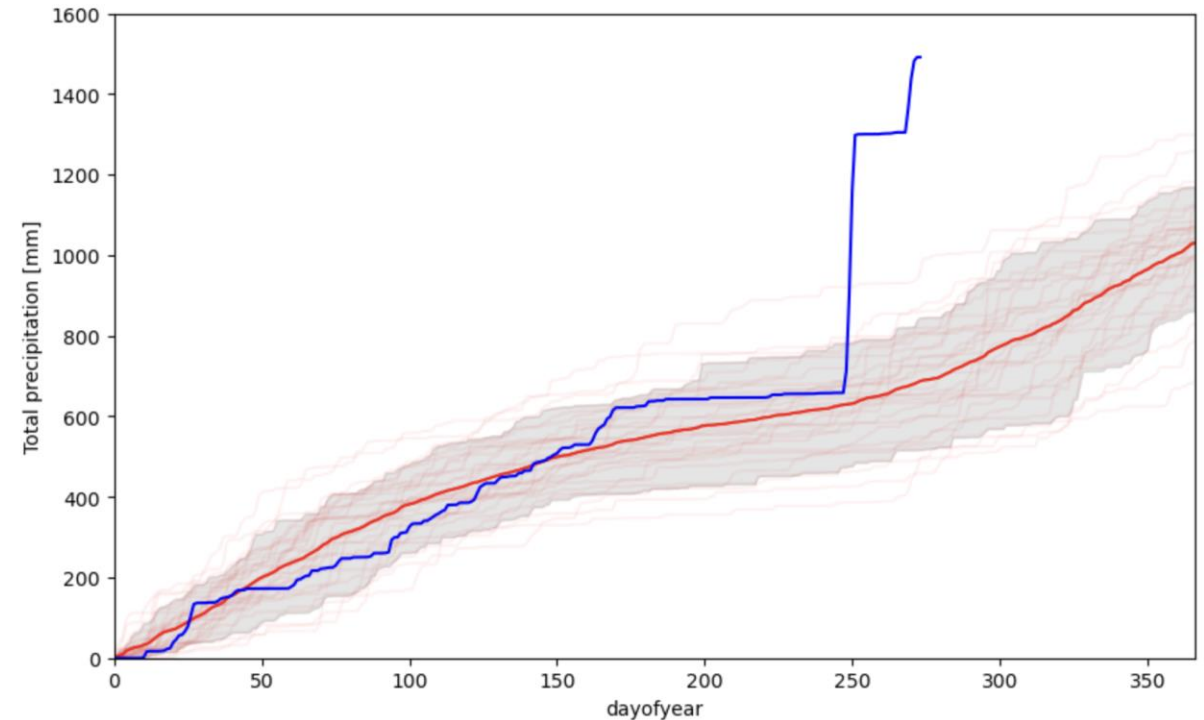
DESP advanced data access

Main catalogue:

- DestinE Climate DT

- DestinE Extremes DT

Supporting datasets:

- C3S: ERA5

- CMEMS: Global Ocean Physics Analysis and Forecast

- Copernicus satellite products

Thank you for your attention