# DSCompare Matlab Package

## Software to analyze, compare and validate analysis and reanalysis datasets with an observed dataset

# USER MANUAL

By
Humberto L. Varona
And
Tonia A. Capuano

Version 1.1

# Software to analyze, compare and validate analysis and reanalysis datasets with an observed dataset (DSCompare).

## *Overview*

Computational tool that analyzes, compares and validates analysis and reanalysis datasets with an observed dataset using statistical tests such as Mann Whitney (U-test), t-test, F-test, Root Mean Square Error (RMSE), correlation coefficient, BIAS, normalized BIAS, trend, scatter index and maximum anomalies.

## *Version*

1.1

## *Release date*

January, 9th 2023

## *License*

MIT

Download URL

https://zenodo.org/record/7517163

## *Cite as*

Varona, Humberto L., Capuano, Tonia A., Noriega, C., & Araujo, M. (2022). Software to analyze, compare and validate analysis and reanalysis datasets with an observed dataset (DSCompare). (1.0). Zenodo. https://doi.org/10.5281/zenodo.7517163

## *How to install*

Matlab 2019b compatible software

- Open Matlab.
- Go to APP tab.
- Click on the "Install App" button.
- Select the DSCompare.mlappinstall file.
- In the Install dialog click on the "Install" button.

## *How to run*

Type in the Matlab command window:

>> DSCompare <Enter>

or find DSCompare in the APP tab of Matlab.

**Operation mode**

Figure 1 shows the main screen of the DSCompare Matlab package, which compares an analysis/reanalysis dataset with a reference dataset (observed data). Only datasets stored in standard NetCDF format and complying with the CF-1.6 convention can be used. Datasets produced by hydro-thermodynamic and regional circulation models, such as ROMS, CROCO and NEMO models have a non-standard NetCDF format, since the reference variables of these are different as longitude and latitude are two-dimensional variables, and depth and time may have different names, e.g., time_counter, level, depthu, depthv, depthw, depth, etc.

All these datasets have to be standardized, that is, they have to be fully compatible with the nco, CDO and ncdump tools as the latter will be used in the preparation of the datasets for use in DSCompare. In the case of ROMS and CROCO models, the ROMSTOOLS (Penven et al., 2003) and CROCOTOOLS packages can be easily adapted for this objective, and for the output of the NEMO model there is a converting tool called fcNEMOtoStd v1.3 (Varona, 2023a) .

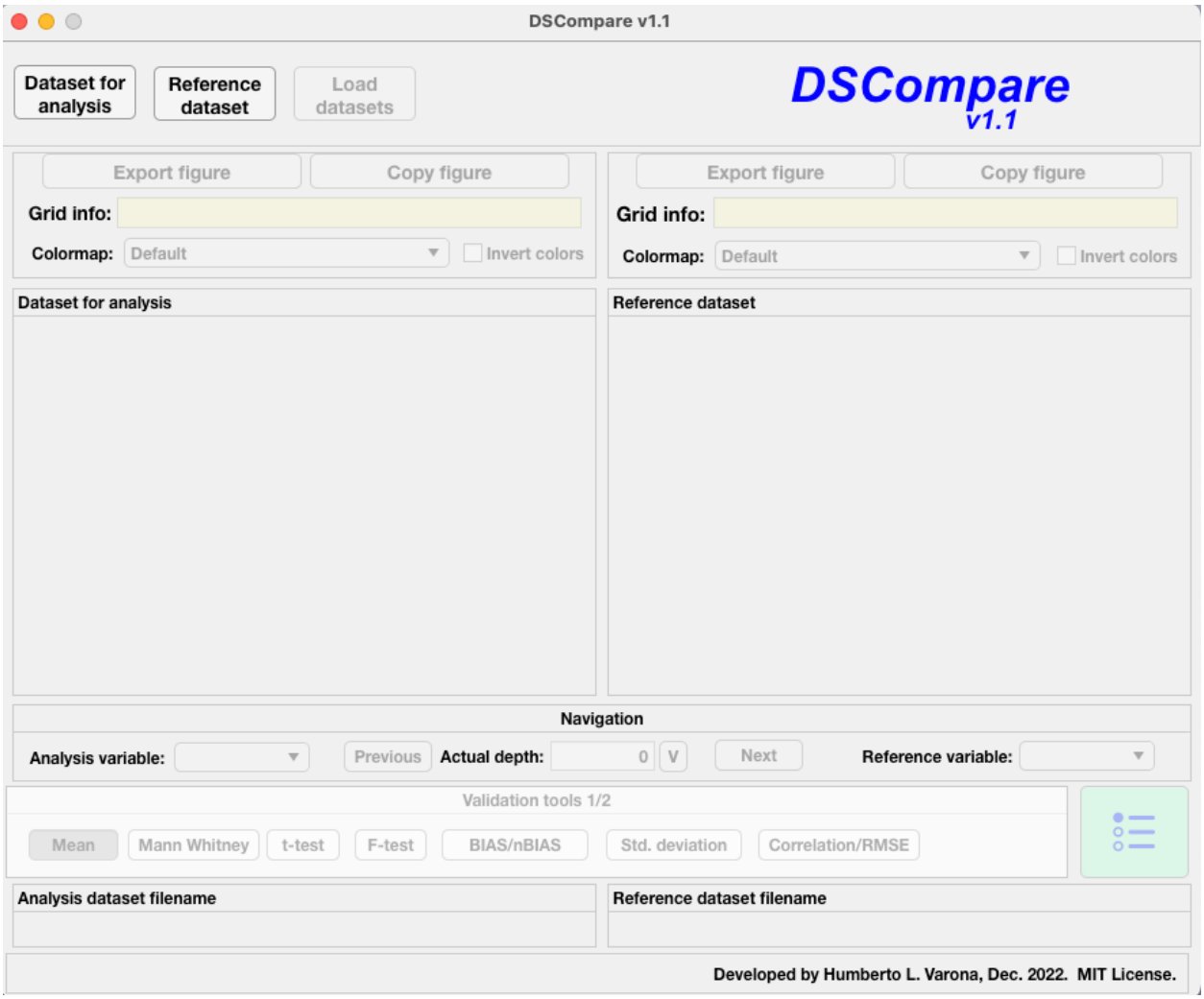

Figure 1. Main screen of DSCompare Matlab package.

## DCompare workflow

1. Select the dataset to be validated by clicking on the "Dataset for analysis" button.
2. Select the variable to be validated using the "Analysis variable" drop-down menu.
3. Select the reference dataset by clicking on the "Reference dataset" button.
4. Select the reference variable using the "Reference variable" drop-down menu.
5. Load both variables into memory using the "Load datasets" button (figure 2).
6. The selected variables will be loaded and all comparison parameters for the surface will be calculated.
7. Through the buttons "Next" and "Previous" it will be possible to change the depth if both datasets have 4 dimensions (lon, lat, depth, time). With the "V" button a specific depth can be selected.
8. Finally, the validations will be performed through the buttons found in "Validation tools"; by default, 7 tools are shown (Mean, Mann Whitney, t-test, F-Test, BIAS/nBIAS, Std.

Deviation and Correlation/RMSE). By clicking on the button , you can obtain 3 more tools (Maximum anomalies, Scatter index/Covariance and Trend).

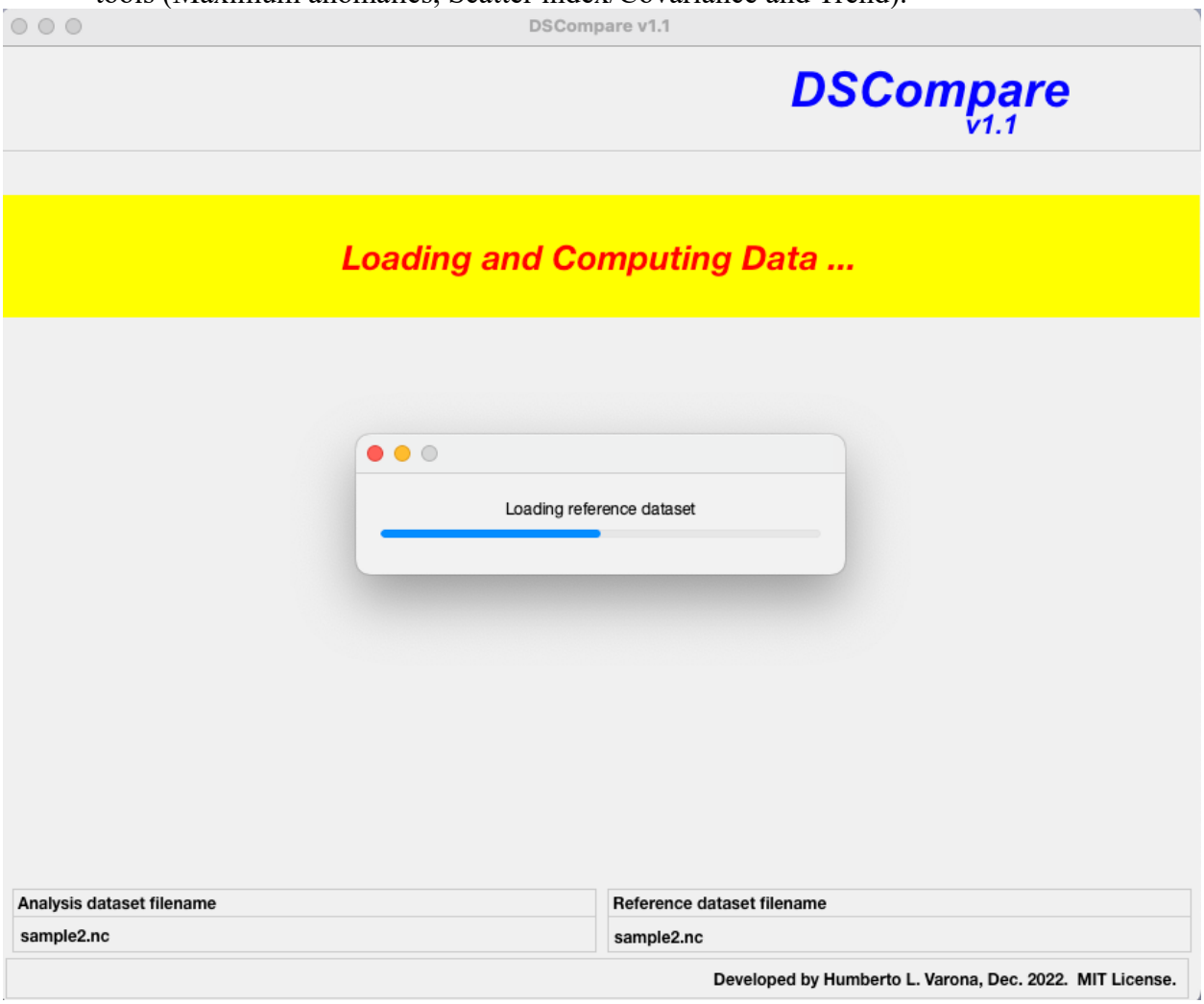


Figure 2. Loading of datasets.

**Preprocessing of the datasets**

The datasets to be analyzed and the reference dataset must have very similar geographical limits; this information can be retrieved using CDO (Schulzweida et al., 2006) from the command line in the terminal:

# cdo sinfo dataset_name.nc

Output:

File format : NetCDF
-1 : Institut Source T Steptype Levels Num Points Num Dtype : Parameter ID
1 : unknown unknown v instant 40 1 6560 1 F64 : -1
2 : unknown unknown v instant 40 1 6560 1 F32 : -2

```
  Grid coordinates :
    1 : lonlat                    : points=6560 (80x82)
                         lon : -59.5 to 19.5 by 1 degrees_east
                         lat : -30.211 to -3.238 by 0.333 degrees_north
  Vertical coordinates :
    1 : pressure                  : levels=40
                        depth : 5 to 4478 millibar
  Time coordinate :
                         time : 489 steps
     RefTime =  0001-01-01 00:00:00  Units = days  Calendar = standard
  YYYY-MM-DD hh:mm:ss  YYYY-MM-DD hh:mm:ss  YYYY-MM-DD hh:mm:ss  YYYY-
MM-DD hh:mm:ss
  1980-01-01 00:00:00  1980-02-01 00:00:00  1980-03-01 00:00:00  1980-04-01 00:00:00
  1980-05-01 00:00:00  1980-06-01 00:00:00  1980-07-01 00:00:00  1980-08-01 00:00:00
.....................................................................................
   ..............................
2020-05-01 00:00:00  2020-06-01 00:00:00  2020-07-01 00:00:00  2020-08-01 00:00:00
  2020-09-01 00:00:00
```

The "sinfo" operator displays the limits of all dimensions of the NetCDF file.

The "sellonlatbox" operator is used to select a geographic region from a dataset:

```
# cdo sellonlatbox,-59.5,19.5,-30.211,-3.238 reference_dataset.nc output.nc
```

The two datasets to be compared have to cover the same temporal period. To select a time interval, type:

```
# cdo seldate, 1980-01-01, 2020-09-01 reference_dataset.nc output.nc
```

They must also match in spatial resolution, both horizontally (this is resolved in DSCompare) and vertically; the latter can be done by means of the "intlevel" operator.

The depths can be displayed as:

```
# ncdump -v depth  reference_dataset.nc

…
…
…
data:

depth = 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 105, 115, 125, 135, 145, 155,
    165, 175, 185, 195, 205, 215, 225, 238, 262, 303, 366, 459, 584, 747,
    949, 1193, 1479, 1807, 2174, 2579, 3016, 3483, 3972, 4478 ;
}
```

```
# ncdump -v depth analysis_dataset.nc

…
…
…
data:

 depth = 5.5, 7.1, 10, 25, 50, 70, 95, 105 ;
}
```

We can then interpolate the dataset to be validated at the depths at which the data is found in the reference dataset with the following operator:

```
# cdo intlevel,5.5,15,25,35,45,55,65,75,8595,105 analysis_dataset.nc ouput.nc
```

Note: The shallowest depth of the reference dataset is 5 m and that of the analysis dataset is 5.5 m, so it cannot be interpolated for the depth of 5 m, and the same happens with the deepest depth. In addition, it cannot be interpolated for depths greater than 105 m, which is the maximum depth of the analysis dataset. Vertical interpolation can also be performed through the Matlab package: NCVerticalInterp (Varona, 2023b).

Once the dataset to be analyzed and the reference dataset have the same spatial geometry and the same data frequency, they are ready to be used in DSCompare.

Note: The size of the datasets that can be loaded into DSCompare will depend on the size of the RAM available on the computer where the Matlab software is run.

Datasets can be reduced by extracting each variable separately through the CDO "selname" operator:

```
# cdo selname,salt analysis_dataset.nc salt_output.nc
```

The CDO user manual can be downloaded from the following URL:

https://code.mpimet.mpg.de/projects/cdo/embedded/cdo.pdf

**Refrences**

Penven, P., Cambon, G., Tan, T. A., Marchesiello, P., & Debreu, L. (2003). ROMSTOOLS user's guide. *Rapport techn., IRD and LPO/UBO, Laboratoire de Physique des Oceans, Universite de Bretagne Occidentale/UFR Sciences*.

Schulzweida, U., Kornblueh, L., & Quast, R. (2006). CDO user's guide. *Climate data operators, Version*, *1*(6), 205-209.

Varona, Humberto L. (2023a). Format converter from NEMO model to NetCDF standard (fcNEMOtoStd) (1.4). Zenodo. https://doi.org/10.5281/zenodo.7519023

Varona, Humberto L. (2023b). Vertically interpolates NetCDF files (NCVerticalInterp) (1.2). Zenodo. https://doi.org/ 10.5281/zenodo.7519015