Community Intercomparison Suite

Short Talk about CISTools at TGIF/DKRZ Meeting

1th March 2018

Based on the Workshop Presentation in 2016 by the CISTools Authors.

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE ENVIRONMENT

Science & Technology Facilities Council

UNIVERSITY OF OXFORD

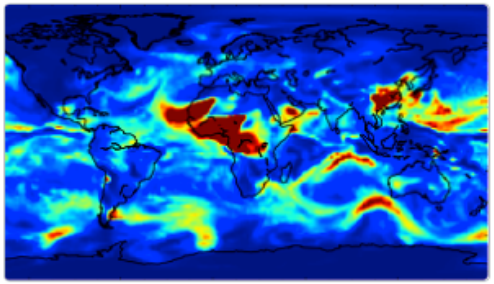
What is CIS?

"CDO for Observational Data": Satellite, in-situ and model.

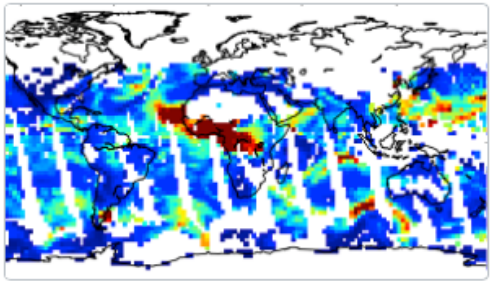
Read

Analyse

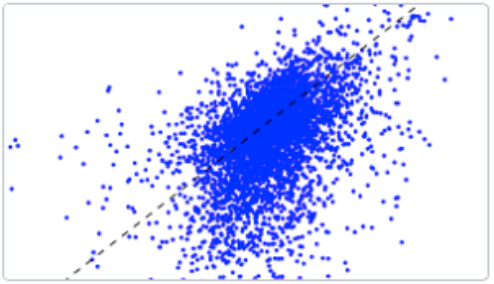
Visualise



Read your data



Analyse your data



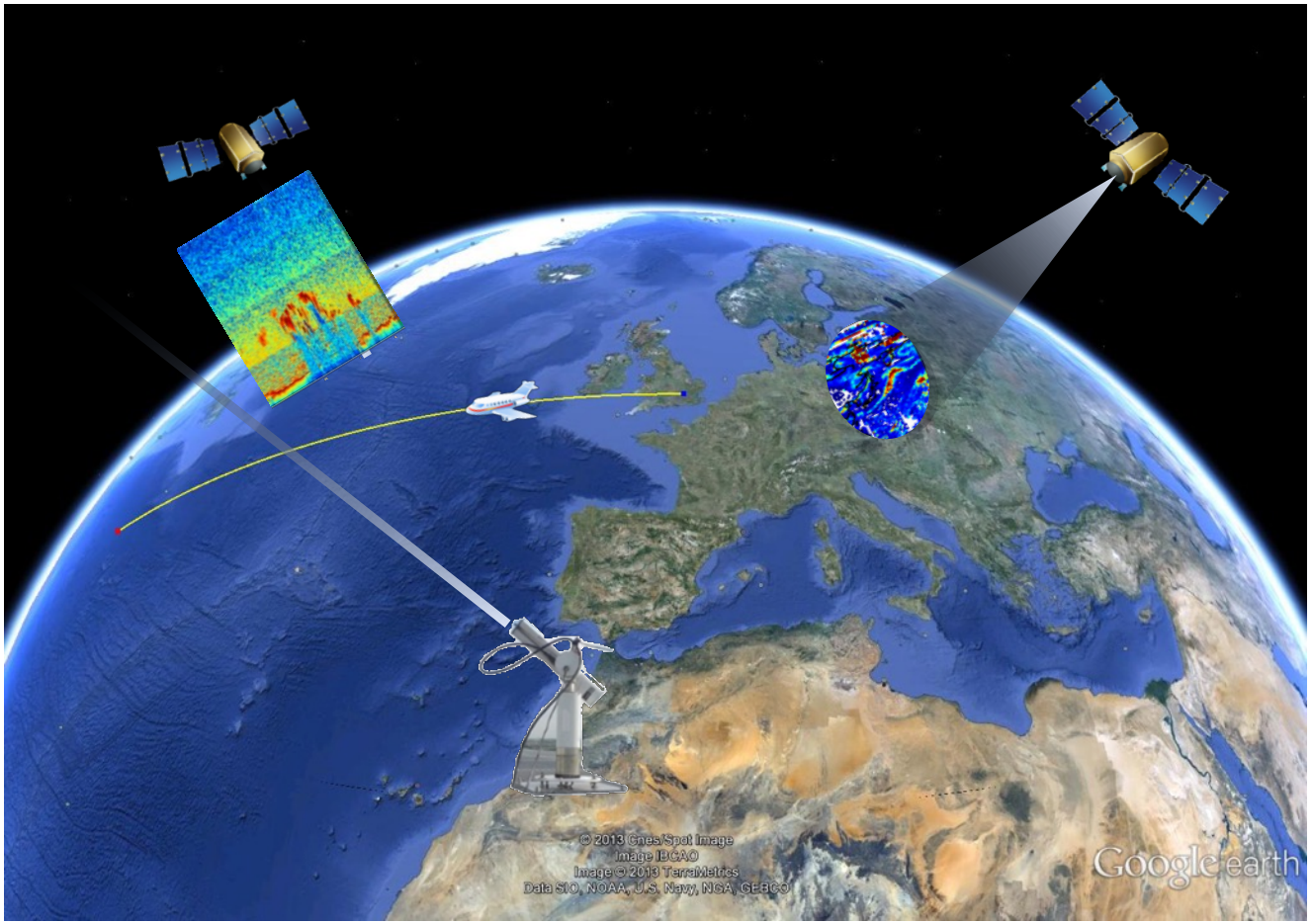
Visualise your data

Extensible data reading

Read

Analyse

Visualise

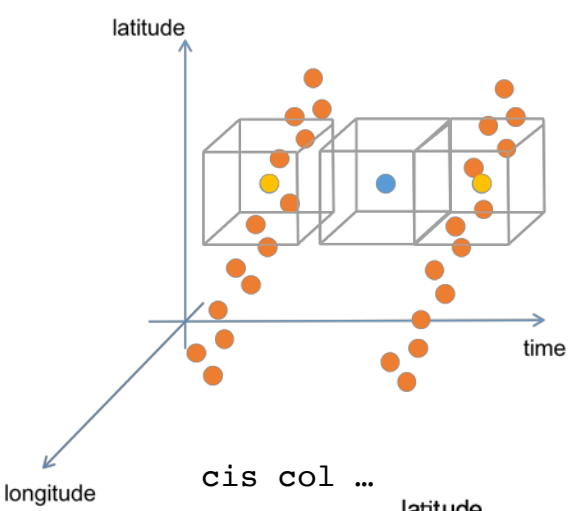


Flexible data analysis

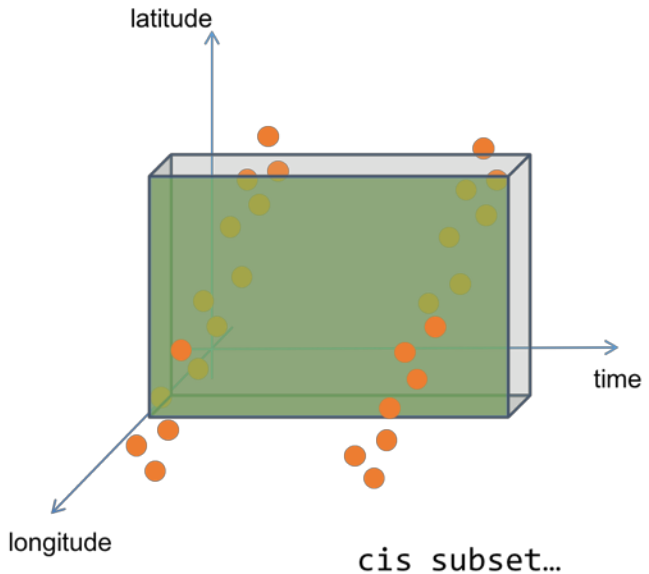
Read

Analyse

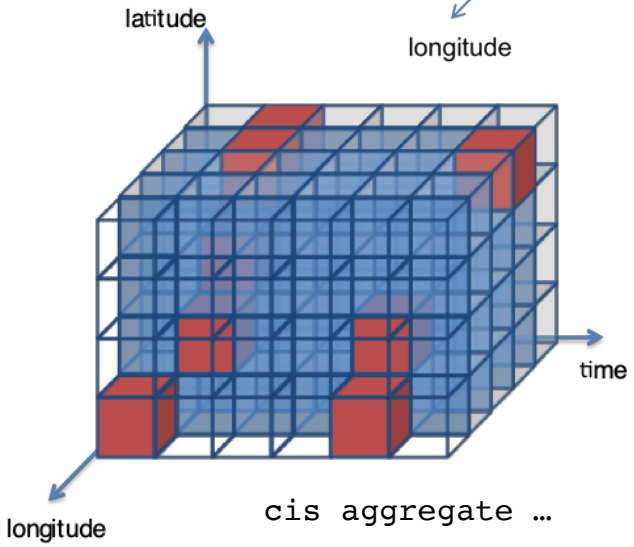
Visualise



`cis col …`



`cis subset…`



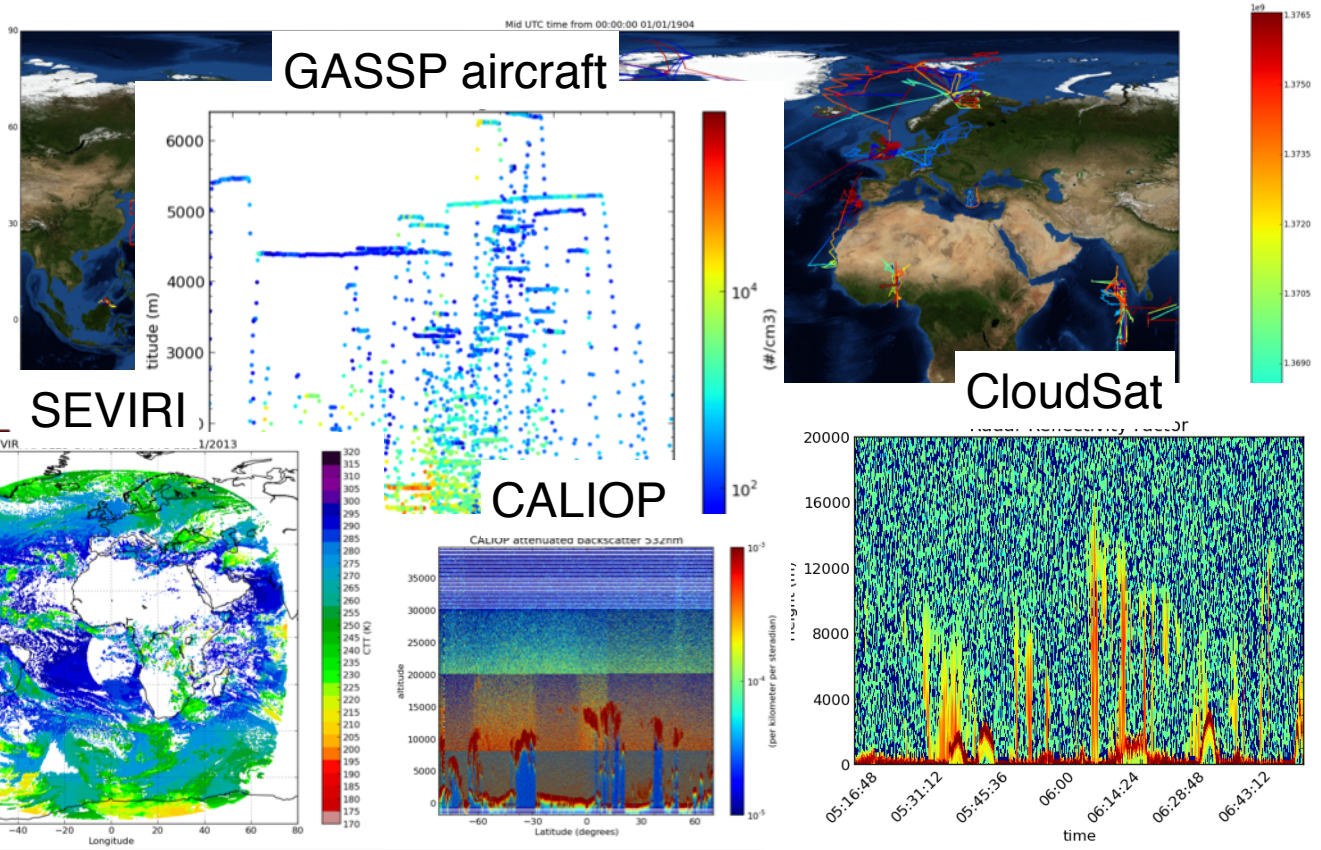
`cis aggregate …`

Comprehensive visualisations

Read

Analyse

```
cis plot…
```



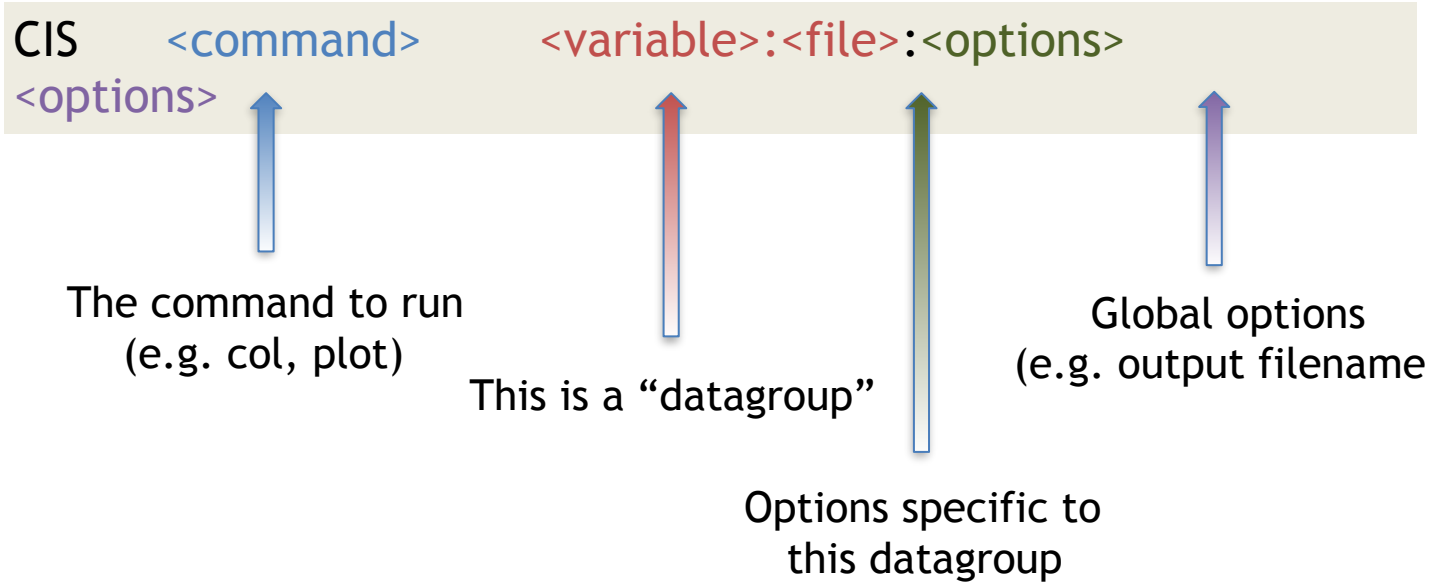
Community Intercomparison Suite

UNIVERSITY OF OXFORD

BASIC INTERFACE INTRODUCTION

Command line interface

- Basic structure



Command line interface

- Basic structure
- Variable wildcarding

```
CIS  aggregate AOD*:<file>:<options>  <options>
```

Command line interface

- Basic structure
- Variable wildcarding
- Filename wildcarding

```
CIS  aggregate “AOD*”:“my_*_file_?_[0-9]”:<options>  <options>
```

Command line interface

- Basic structure
- Variable wildcarding
- Filename wildcarding
- Much more detailed help can be found at: cis.readthedocs.org

HANDS-ON WITH CIS

Hands-on with CIS

- Install cistools with conda:

```
conda create -c conda-forge -n cis_env cis

source activate cis_env
```

- And download the test data
- Then you can use CIS to get a feel for the contents of the files:

```
cis  info ARCPAC_2008
```

Lat/lon plots

Similarly we can plot the aircraft data:

```
cis plot NUMBER_CONCENTRATION:ARCPAC_2008
```

Or, perhaps adding a nice background:

```
cis plot NUMBER_CONCENTRATION:ARCPAC_2008 --nasabluemarble
```

We can plot the satellite data too:

```
cis plot AOD550:AerosolCCI/20080415*.nc --xmin -170 --xmax -100 --ymin 35 --yma
```

Global model plot

Show the variables in a CMIP5 data file:

```
cis  info cmip5/tas.nc
```

Taking a time average:

```
cis  aggregate tas:cmip5/tas.nc t  -o tas_average.nc
```

And plot it:

```
cis  plot  tas:tas_average.nc
```

Take a subset and plot it again:

```
cis  subset  tas:tas_average.nc  x=[-0,50],y=[30,80]  --o tas_subset.nc
cis  plot  tas:tas_subset.nc
```

Comments from the Author

- CDO for observational data
- started as a command line tool but is mostly used as a Python library now
- builds on Iris (MetOffice)
- main selling point is reading weird and wonderful satellite and in-situ (and model!) data into an Iris Cube-like object, and allowing easy spatio-temporal collocation between them

Community and Future Plans?

- There are lots of downloads (20k+) and it appears to have a healthy user base
- Currently only me developing it in my spare time
- Our future plans are mostly around making the most of dask. Moving to Dask/xarray with version 2.0 in 2019.
- In touch with the ESMValTool team about using CIS for reading observational data directly into their analysis code

Summary

- CIS is an open source python toolbox for reading, analysing and visualising earth sciences data
- Lots of support for community developed plugins
- There is also a Python API available
- Future work:
  - Support for ‘hybrid’ semi-gridded data types
  - Vector plots… and more!

Join us at cistools.net

Follow us @cistoolsnet

Resources

- Homepage: http://cistools.net/
- CIS Workshop in 2016
- Jupyter Notebook
- Example Data
- CIS Paper

SPARE SLIDES

Collocation command

```
CIS col <source> <sample>:<options> -o <output>
```

Data that will be remapped

New spatio-temporal sampling

```
CIS col <variable>:<model data> <model data>:collocator=lin
CIS col <variable>:<model data> <obs data>:collocator=lin
CIS col <variable>:<obs data> <model data>:collocator=bin,kernel=mean
CIS col <variable>:<obs data> <other obs data>:collocator=box[<options>]
```

Collocation options

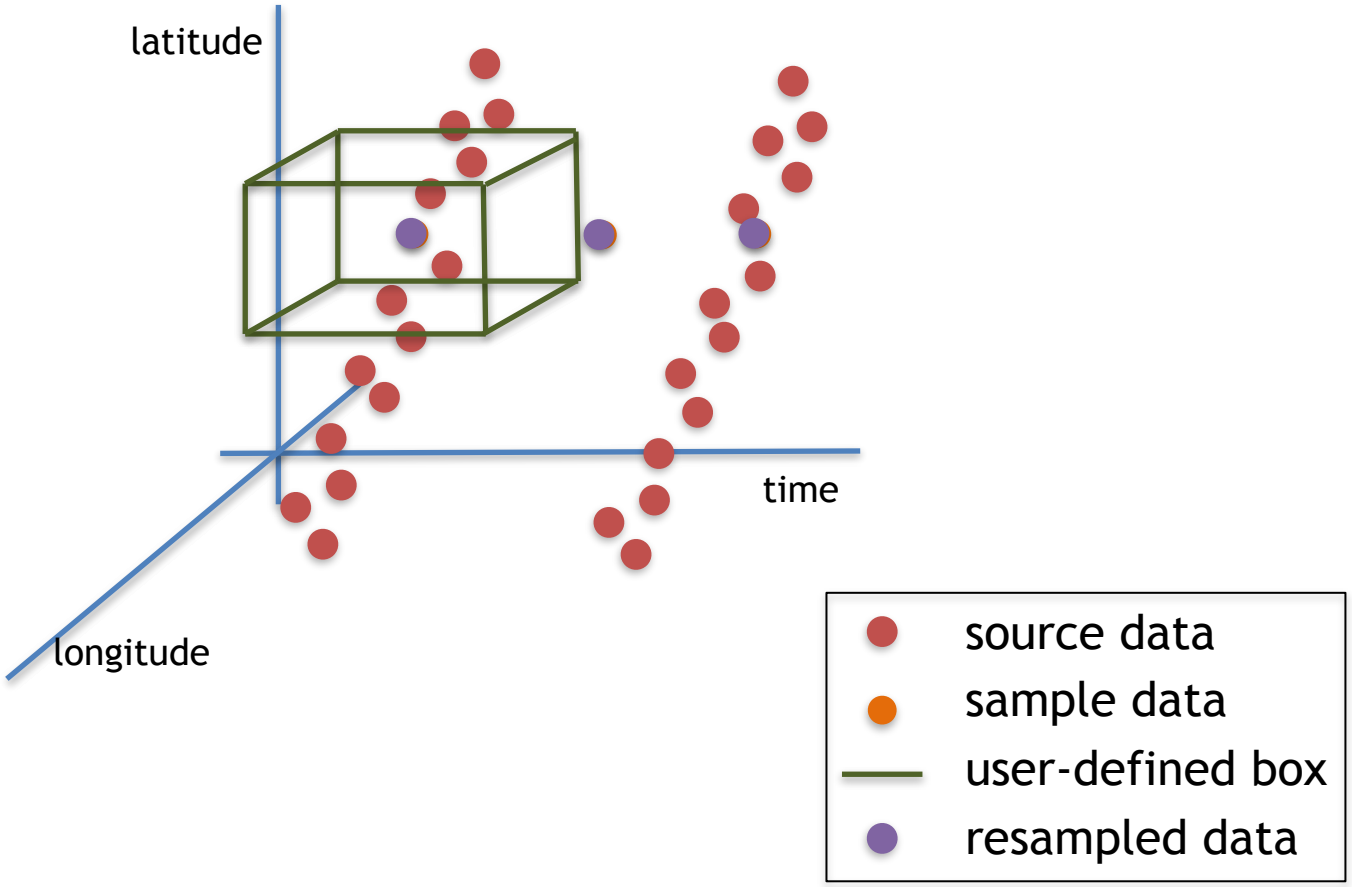
sample

source

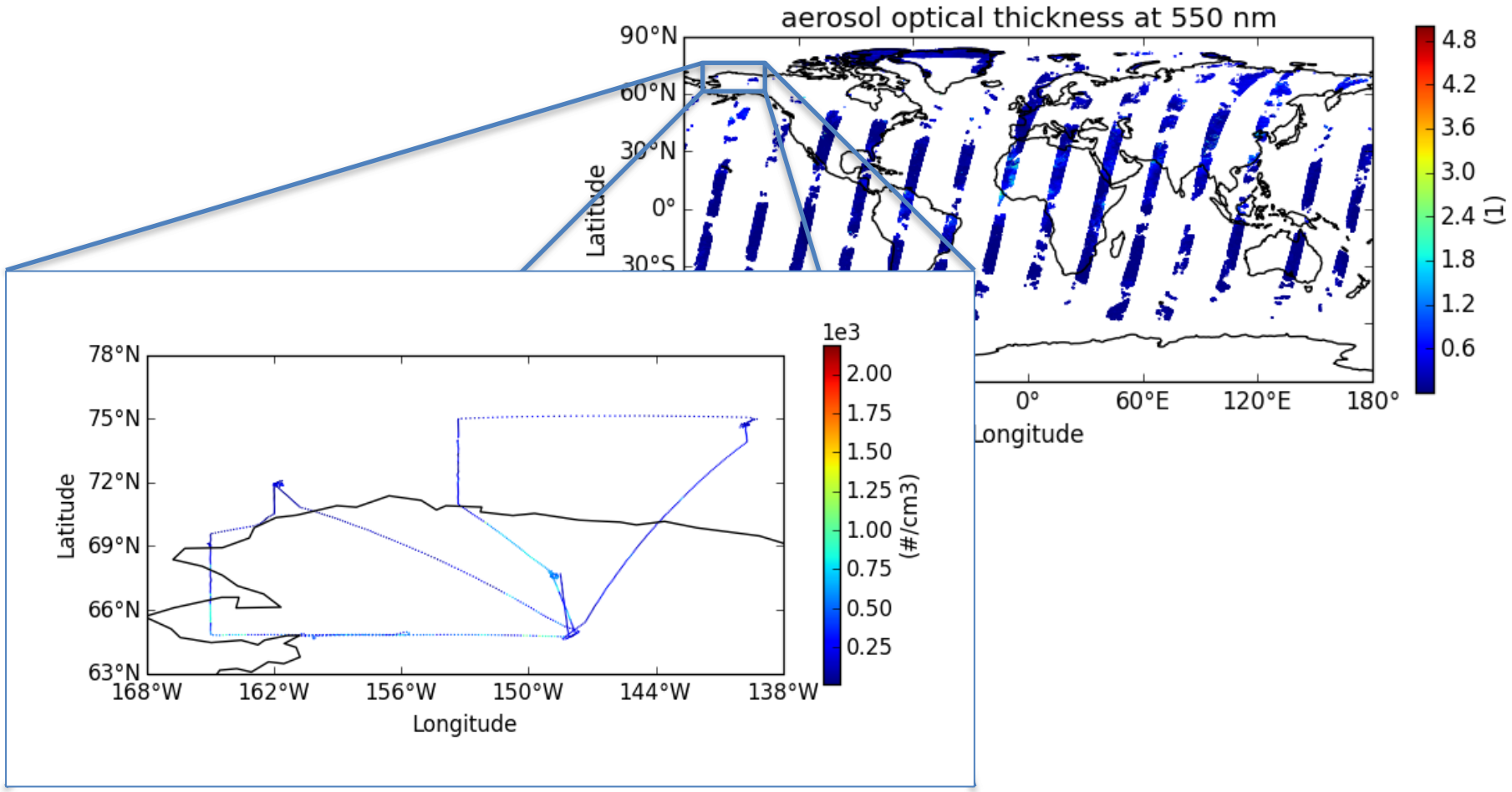
	gridded	ungridded
gridded	***lin***, nn, box	lin, ***nn***
ungridded	box, ***bin***	***box***

lin : linear interpolation in space and time
nn : Nearest Neighbour
bin : operates *kernel* on all *source* data in *sample* bounds
box : operates *kernel* on all *source* data in user-defined box on *sample*

bin/box use in collocation



Aircraft onto Satellite



Aircraft onto Satellite

First subset the satellite data:

```
CIS subset AOD*:AerosolCCI x=[-170,-100],y=[35,80] --o aerosol_cci_subset_Alask
```

We now have a much reduced dataset:

```
CIS plot AOD550:cis-aerosol_cci_subset_Alaska.nc --itemwidth=4
```

Now we do the collocation:

```
CIS col NUMBER_CONCENTRATION,SUPERSATURATION:ARCPAC_2008/*.nc
        cis-
aerosol_cci_subset_US.nc:variable=AOD550,collocator=box[h_sep=10,t_sep=P1D]
        -o collocated_ccn_AOD550
```

Aircraft onto Satellite

Now we can do a comparison plot:

```
CIS plot NUMBER_CONCENTRATION:cis-ccn_collocated_to_AOD550.nc
         AOD550:cis-aerosol_cci_subset_Alaska.nc
         --type comparativescatter --itemwidth=4
```

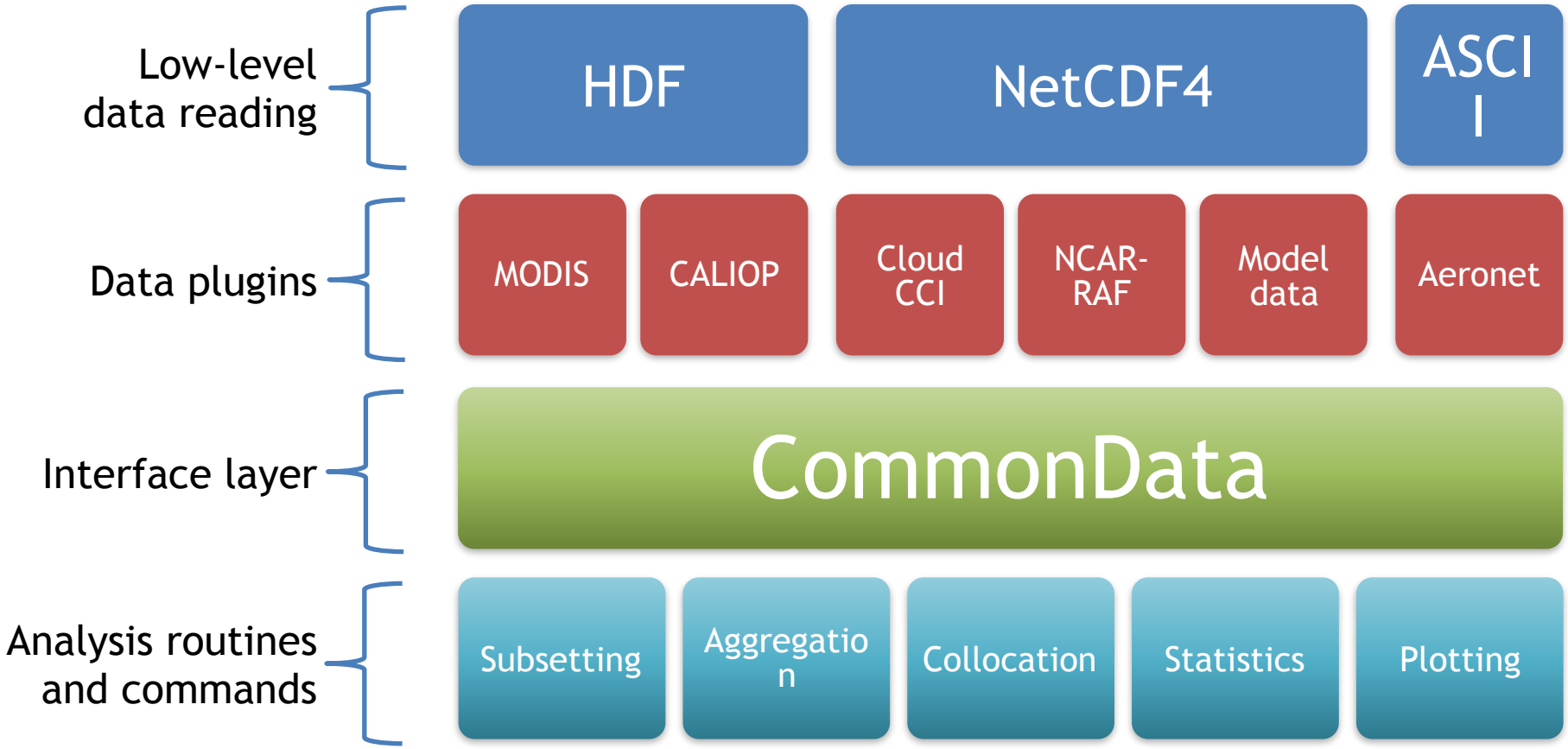
Or just plot the collocated data:

```
CIS plot NUMBER_CONCENTRATION:cis-ccn_collocated_to_AOD550.nc --itemwidth=
```

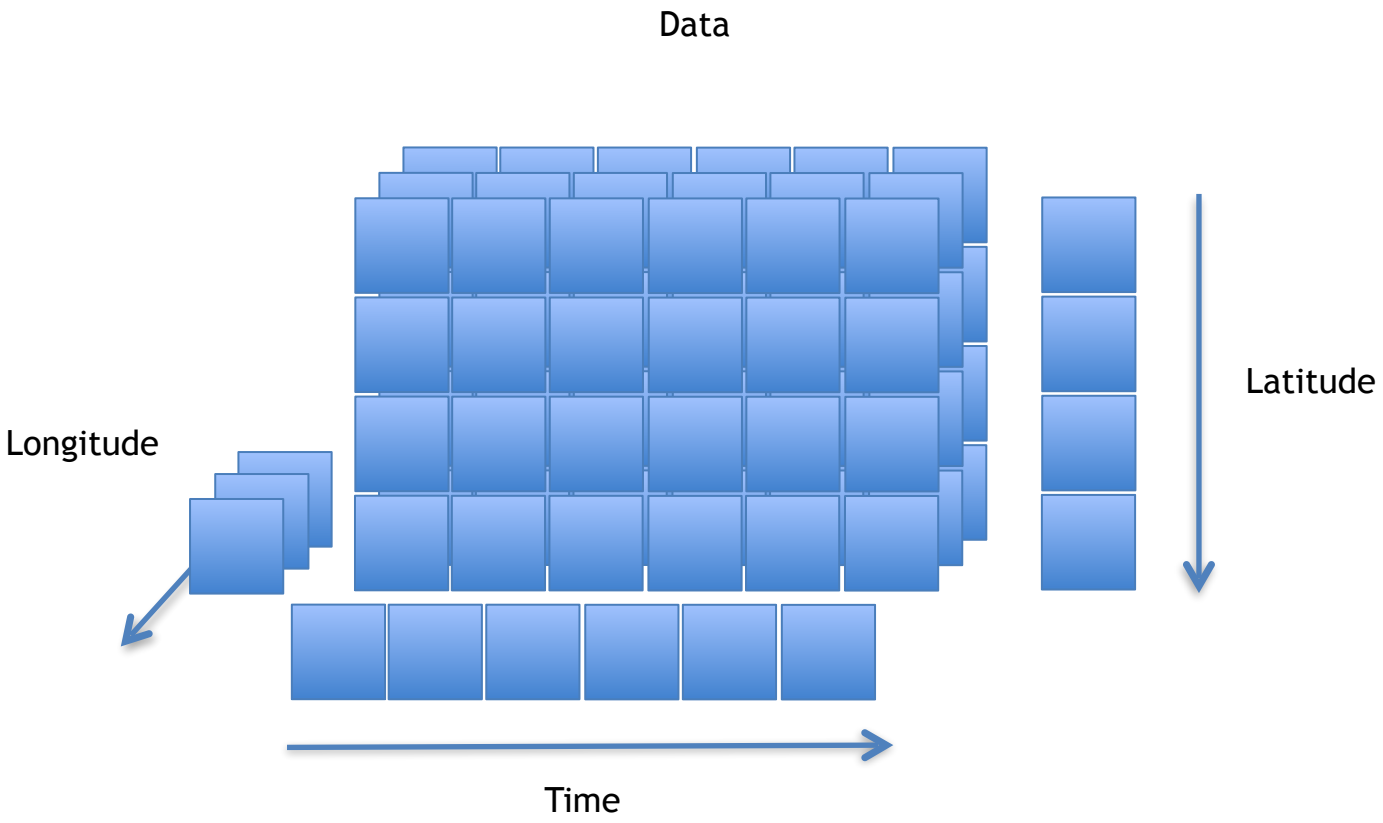
Lets look at the correlation for these few points:

```
CIS stats NUMBER_CONCENTRATION:cis-collocated_ccn_AOD550.nc
          AOD550:cis-aerosol_cci_subset_US.nc
```

Architecture



GriddedData



UngriddedData