Python and ecCodes

Xavi Abellan

Xavier.Abellan@ecmwf.int

Python and GRIB API

- Just an appetizer
- Provide you only a small view of the world the Python interface opens to
- Increase your awareness
- You need to explore!

What is Python?

- Interpreted, high level scripting language
- Strong, but optional, Object Oriented programming support
- Open-source software, which means it's free
- Easy to learn
- Portable
- Dynamic typing
- Support for exception handling
- Good integration with other languages
- Higher productivity
- Alternative to Matlab, IDL, …
- Through extensions supports many scientific data formats, e.g. netcdf, hdf5, grib, etc.

Python basics: hello world

- Import the modules you need
- Indentation to define the different blocks:
  - No ; or { } or END
- Function definition with def
- Variable types not explicitly defined
- Dealing with strings is easy…
- Run with python or directly if shebang present and permissions set

```
#!/usr/bin/env python
import sys

# This is a comment
def say_hello(name):
    print(“Hello "+ name + "!" )

if len(sys.argv) > 1 :
    name = sys.argv[1]
else:
    name = “World"

say_hello(name)
```

```
$> python example.py
hello World!
$> ./example.py Xavi
hello Xavi!
```

Python basics: list and dicts

```
$> python
>>> mylist = ['a','b','c']
>>> print(mylist)
['a', 'b', 'c']
>>> mylist[2:]
['c']
>>> mylist[-1]
'c'
>>> for element in mylist:
...     print(element)
...
a
b
c
```

```
>>> mydict = {'key1':1,'key2':2,'key3':3}
>>> for key,value in mydict.items():
...     print(key + ":" + str(value))
...
key3:3
key2:2
key1:1
>>> 'key1' in mydict
True
>>> 'key5' in mydict
False
>>> len(mydict)
3
>>> mydict.keys()
['key3', 'key2', 'key1']
>>> mydict.values()
[3, 2, 1]
```

NumPy

- Fundamental Python package for scientific computing
- Provides support for multidimensional arrays
- Good assortment of routines for fast operations on arrays
- Performance comparable to that of C or Fortran
- A growing number of Python-based mathematical and scientific packages are using NumPy
- At its core is the ndarray object, an n-dimensional array of homogenous data

```
>>> from numpy import *
>>> a = arange(15).reshape(3, 5)
>>> a
array([[ 0,  1,  2,  3,  4],
       [ 5,  6,  7,  8,  9],
       [10, 11, 12, 13, 14]])
>>> a.shape
(3, 5)
>>> a.ndim
2
>>> a.size
15
>>> b = array([6, 7, 8])
>>> b
array([6, 7, 8])
>>> a.sum()
105
>>> a.min()
0
>>> a.max()
14
>>> a.mean()
7.0
>>> b*2
array([12, 14, 16])
>>> b-b
array([0, 0, 0])
>>> b*b
array([36, 49, 64])
```

NumPy

"""It can be hard to know what functions are available in NumPy."""

http://docs.scipy.org/doc/numpy/reference/

- Operations on arrays:
  - Mathematical and logical
  - Shape manipulation
  - Selection
  - I/O
  - Discrete Fourier transforms
  - Basic linear algebra
  - Basic statistical functions
  - Random simulation

matplotlib

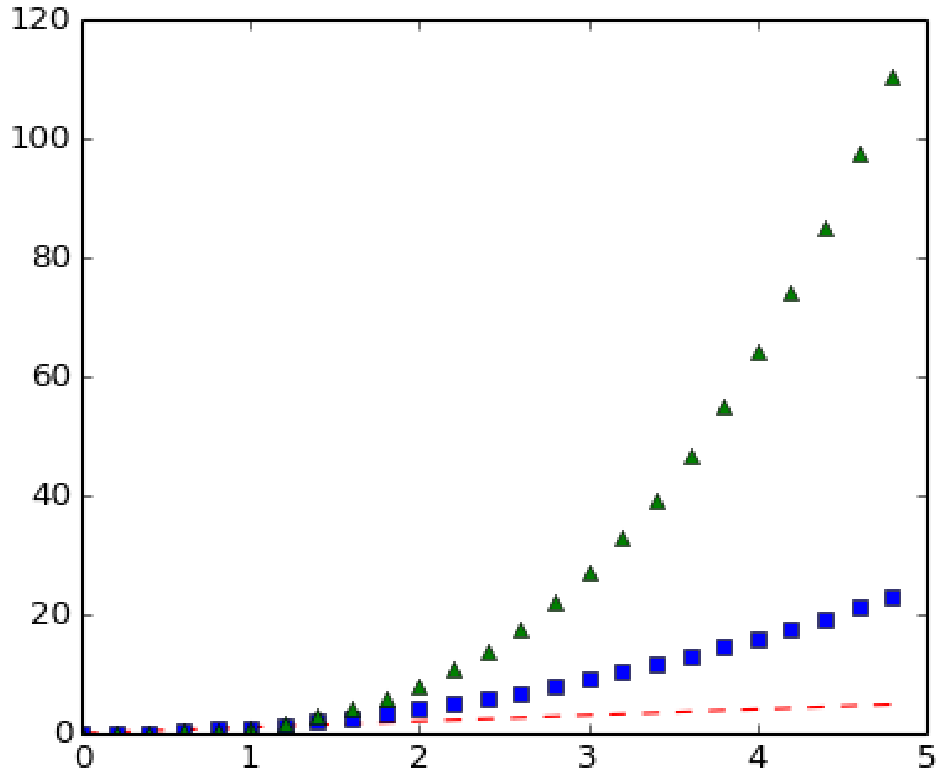
- Plotting library for Python and Numpy extensions
- Has its origins in emulating the MATLAB graphics commands, but it is independent
- Uses NumPy heavily
- Its philosophy is:
  - It should be easy to create plots
  - Plots should look nice
  - Use as few commands as possible to create plots
  - The code used should be easy to understand
  - It should be easy to extend code
- Supports 2D and 3D plotting
- Basemap module: projections, coastlines, political boundaries

matplotlib

```
import numpy as np
import matplotlib.pyplot as plt

# evenly sampled time at 200ms intervals
t = np.arange(0., 5., 0.2)

# red dashes, blue squares and green triangles
plt.plot(t, t, 'r--', t, t**2, 'bs', t, t**3, 'g^')
plt.show()
```



matplotlib

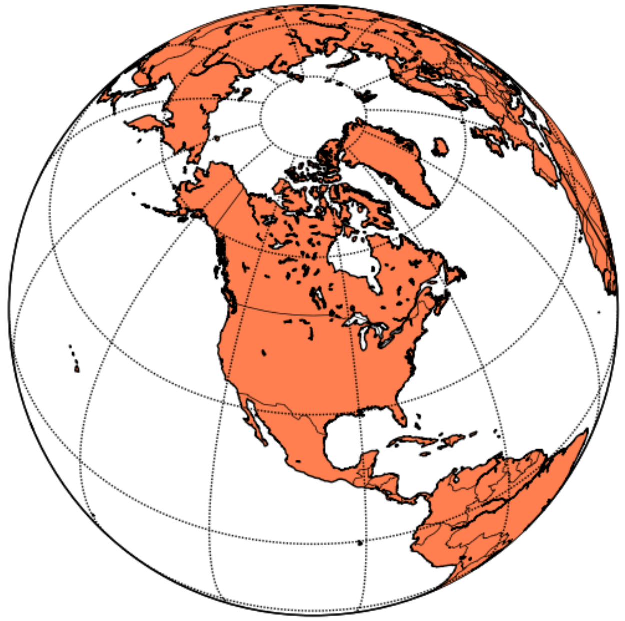
```
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import numpy as np

# make sure the value of resolution is a lowercase L,
#  for 'low', not a numeral 1
map = Basemap(projection='ortho', lat_0=50, lon_0=-100,
              resolution='l', area_thresh=1000.0)

map.drawcoastlines()
map.drawcountries()
map.fillcontinents(color='coral')
map.drawmapboundary()

map.drawmeridians(np.arange(0, 360, 30))
map.drawparallels(np.arange(-90, 90, 30))

plt.show()
```

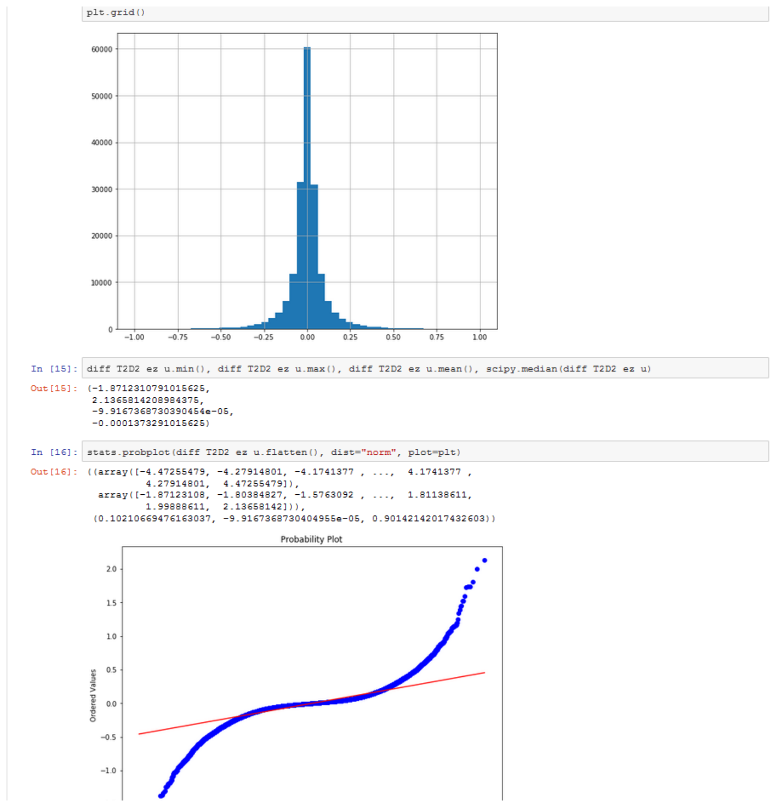


SciPy library

- Open source library of scientific algorithms and mathematical tools
- Dependent on NumPy
- Offers improved versions of many NumPy functions
- Quite fast as most of its calculations are implemented in C extension modules
- Offers a decent selection of high level science and engineering modules for:
  - statistics
  - optimization
  - numerical integration
  - linear algebra
  - Fourier transforms
  - signal processing
  - image processing
  - ODE solvers
  - special functions

Ipython – Jupyter

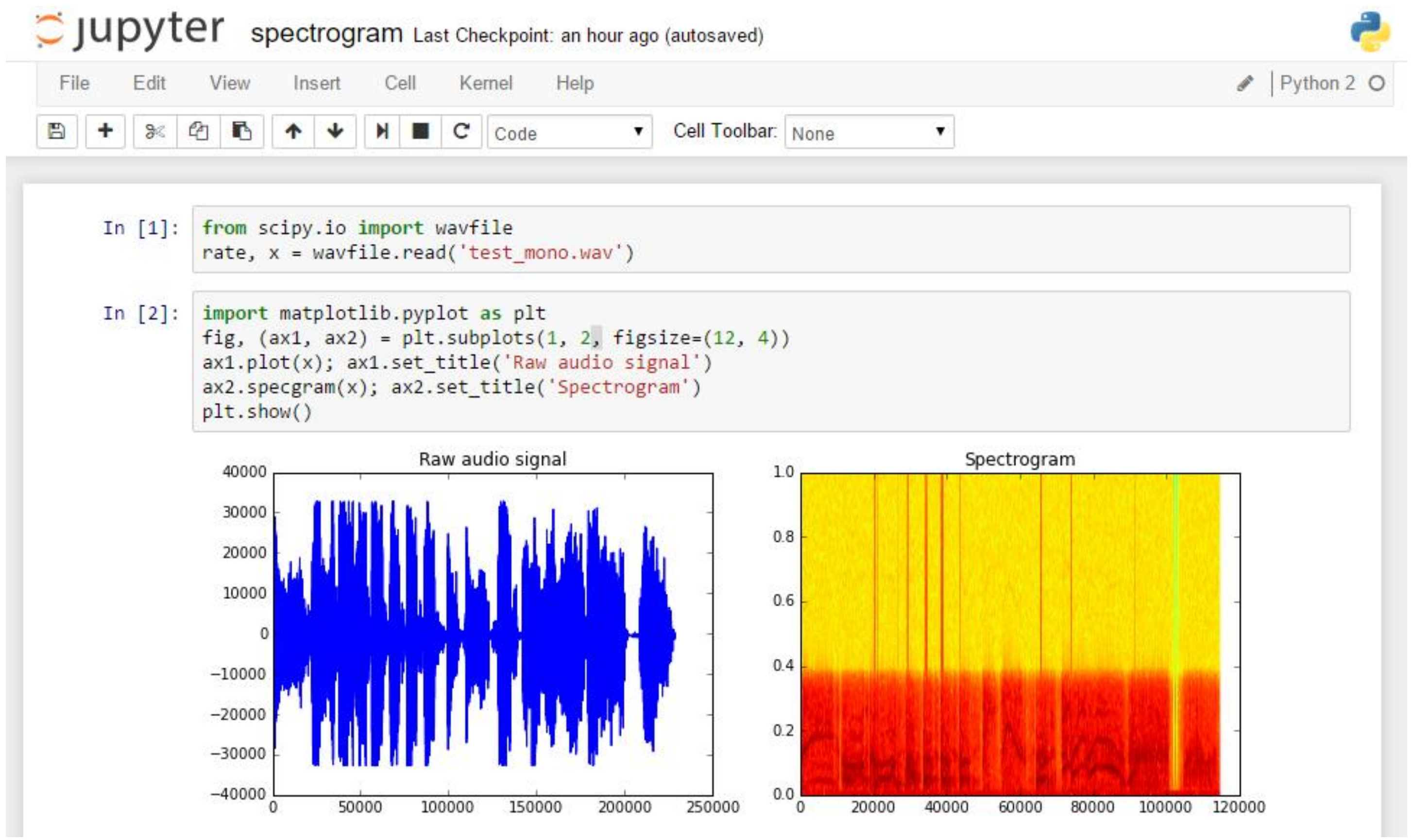
- Interactive and enhanced python console
- Server/client web Notebooks



```
plt.grid()
```

```
In [15]: diff T2D2 ez u.min(), diff T2D2 ez u.max(), diff T2D2 ez u.mean(), scipy.median(diff T2D2 ez u)
Out[15]: (-1.8712310791015625,
          2.1365814208984375,
          -9.9167368730390454e-05,
          -0.0001373291015625)

In [16]: stats.probplot(diff T2D2 ez u.flatten(), dist="norm", plot=plt)
Out[16]: ((array([-4.47255479, -4.27914801, -4.1741377 , ...,  4.1741377 ,
                   4.27914801,  4.47255479]),
           array([-1.87123108, -1.80384827, -1.5763092 , ...,  1.81138611,
                   1.99888611,  2.13658142])),
          (0.10210669476163037, -9.9167368730404955e-05, 0.90142142017432603))
```



```
In [1]: from scipy.io import wavfile
        rate, x = wavfile.read('test_mono.wav')

In [2]: import matplotlib.pyplot as plt
        fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
        ax1.plot(x); ax1.set_title('Raw audio signal')
        ax2.specgram(x); ax2.set_title('Spectrogram')
        plt.show()
```

SciPy Software stack

- Python-based ecosystem of open-source software for mathematics, science, and engineering
- It depends on other python packages like:
  - **Numpy**: Base N-dimensional array package
  - **SciPy library** : Fundamental library for scientific computing
  - **Matplotlib**: Comprehensive 2D Plotting
  - **Ipython / Jupyter**: Enhanced Interactive Console, notebooks
  - **Sympy**: Symbolic mathematics
  - **Pandas**: Data structures & analysis

Python at ECMWF

- Currently two interfaces for ECMWF libraries
  - ecCodes / GRIB API
  - Magics++
- ecCharts
- New web plots (ecCodes, magics++)
- Verification (ecCodes, magics++)
- EcFlow (SMS's replacement) - server configuration and client communication
- Copernicus Project (ecCodes)
- EFAS (European Flood Alert System) (EcFlow)
- Research
- Python interface for future interpolation library is planned

Magics++

- ECMWF's inhouse meteorological plotting software
- Used at ECMWF and in the member states for more than 25 years
- Supports the plotting of contours, wind fields, observations, satellite images, symbols, text, axis and graphs
- Two different ways of plotting
  - Data formats which can be plotted directly: GRIB1, GRIB2, BUFR, ODB, NetCDF and NumPy
  - Data fields can be read with ecCodes, can be modified and then passed to magics++ for plotting
- The produced meteorological plots can be saved in various formats, such as PS, EPS, PDF, GIF, PNG, KML and SVG
- Provides both a procedural and a high-level Python programming interface

Python in ecCodes and GRIB API

- Available since GRIB API version 1.9.5
- Python 2.5 or higher required. Python 3 not yet supported
- Low level, procedural
- Provides almost 1 to 1 mappings to the C API functions
- Uses the NumPy module natively to handle data values
- Should be available at ECMWF through module system
  - Use module to change the version

Python API – Enabling

- If building the library by hand:

```
cmake –DENABLE_PYTHON ..
```

- On 'make install', the Python API related files will go to:

```
{prefix}/lib/pythonX.X/site-packages/eccodes
{prefix}/lib/pythonX.X/site-packages/gribapi
```

- Either set the PYTHONPATH or link to these files from your Python
- Ready to go:

```
import eccodes
import gribapi
```

Python API – Loading/Releasing a GRIB message

gid = ***codes_grib_new_from_file*** (file, headers_only=False) ***codes_any_new_from_file*** ***codes_new_from_file*** (file, **product_kind**, headers_only) - CODES_PRODUCT_GRIB - CODES_PRODUCT_BUFR - CODES_PRODUCT_ANY	***grib_new_from_file*** Returns a handle to a GRIB message in a file. Requires the input file to be a Python file object. The use of the headers_only option is not recommended at the moment.
gid = ***codes_new_from_samples*** (samplename) Returns a handle to a message contained in the samples directory	***grib_new_from_samples***
gid = ***codes_new_from_message*** (message) Returns a handle to a message in memory	***grib_new_from_message***
codes_release (gid) Releases the handle	***grib_release***

Python API – Decoding

value = ***codes_get*** (gid, key, ktype=None) Returns the value of the requested key in the message gid is pointing to in its native format. Alternatively, one could choose what format to return the value in (int, str or float) by using the type keyword.	***grib_get***
values = ***codes_get_array*** (gid, key, ktype=None) Returns the contents of an array key as a NumPy ndarray or Python array. type can only be int or float.	***grib_get_array***
values = ***codes_get_values*** (gid) Gets data values as 1D array	***grib_get_values***
On error, a ***CodesInternalError*** exception (which wraps errors coming from the C API) is thrown	***GribInternalError***

Python API – Utilities

[outlat, outlon, value, distance, index] = ***codes_grib_find_nearest*** (gid, inlat, inlon, is_lsm=False, npoints=1) Find the nearest point for a given lat/lon With npoints=4 it returns a list of the 4 nearest points	***codes_find_nearest***
iter_id = ***codes_grib_iterator_new*** (gid,mode)	***grib_iterator_new***
[lat,lon,value] = ***codes_grib_iterator_next*** (iterid)	***grib_iterator_next***
codes_grib_iterator_delete (iter_id)	***grib_iterator_delete***

Python API – Indexing

iid = ***codes_index_new_from_file*** (file, keys) Returns a handle to the created index	***grib_index_new_from_file***
codes_index_add_file (iid, file) Adds a file to an index.	***grib_index_add_file***
codes_index_write (iid, file) Writes an index to a file for later reuse.	***grib_index_write***
iid = ***codes_index_read*** (file) Loads an index saved with *codes_index_write* to a file.	***grib_index_read***
codes_index_release (iid) Release the index	***grib_index_release***

Python API – Indexing

size = ***codes_index_get_size*** (iid, key) Gets the number of distinct values for the index key.	***grib_index_get_size***
values = ***codes_index_get*** (iid, key, ktype=str) Gets the distinct values of an index key.	***grib_index_get***
codes_index_select (iid, key, value) Selects the message subset with key==value.	***grib_index_select***
gid = ***codes_new_from_index*** (iid) Same as *codes_grib_new_from_file* Release with *codes_release*(gid)	***grib_new_from_index***

Python API – Encoding

codes_set (gid, key, value) Sets the value for a scalar key in a grib message.	***grib_set***
codes_set_array (gid, key, value) Sets the value for an array key in a grib message. The input array can be a numpy.ndarray or a Python sequence like tuple, list, array, ...	***grib_set_array***
codes_set_values (gid, values) Utility function to set the contents of the 'values' key.	***grib_set_values***
clone_id = ***codes_clone*** (gid_src) Creates a copy of a message. You can directly write to file with *codes_write* Don't forget to *codes_release*	***grib_clone***

Python API – Exception handling

- All ecCodes functions throw the following exception on error:

CodesInternalError

- All GRIB API functions throw the following exception on error:

GribInternalError

- Wraps errors coming from the C API

Python API – High Level interface (EXPERIMENTAL)

- High-level, more *pythonic* interface

```
with GribFile(filename) as grib:
    # Iterate through each message in the file
    for msg in grib:
        # Access a key from each message
        print(msg[key_name])
        # Report message size in bytes
        msg.size()
        # Report keys in message
        msg.keys()
        # Set scalar value
        msg[scalar_key] = 5
        # Array values are set transparently
        msg[array_key] = [1, 2, 3]
        # Messages can be written to file
        with open(testfile, "w") as test:
            msg.write(test)
        # Messages can be cloned from other messages
        msg2 = GribMessage(clone=msg)
```

Example scripts

- ecCodes:
  - index.py: example on indexed access
  - reading.py: example on matplotlib usage
  - geo.py: example on iterating over the lat/lon values
- basemap: example of basemap plotting data from a grib
  - 2t.py, sst.py
- magics: example of plotting using Magics++
  - basic_gribapi.py, basic_magics.py colour_gribapi.py magics.py
- performance: little example comparing the performance of the tool, the Fortran and the python APIs

```
$> cd $SCRATCH
$> tar xvzf ~trx/ecCodes/python-grib-practicals.tar.gz
```

References

Python specifics

http://www.python.org/

NumPy

http://numpy.scipy.org/

http://www.scipy.org/Numpy_Functions_by_Category

http://docs.scipy.org/numpy/docs/numpy/

http://www.scipy.org/NumPy_for_Matlab_Users

Langtangen, Hans Petter, ”Python scripting for computational science”

References

SciPy

http://www.scipy.org/

Matplotlib & Basemap

http://matplotlib.sourceforge.net/

http://matplotlib.org/basemap

ecCodes

https://software.ecmwf.int/wiki/display/ECC/ecCodes+Home

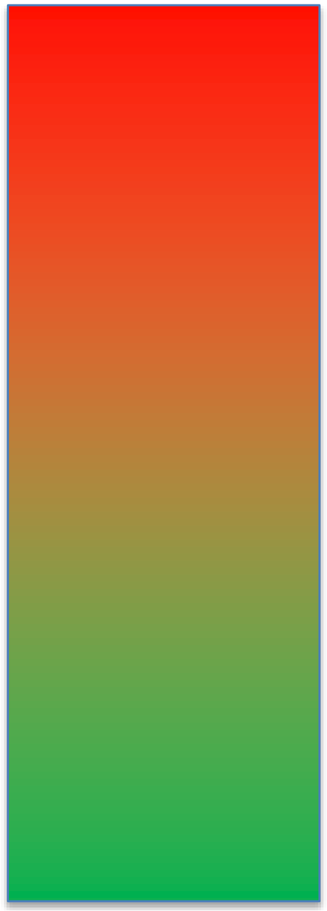
Magics

https://software.ecmwf.int/wiki/display/MAGP

Questions?

THE CHALLENGE

Compute and plot wind speed out of u and v fields. Use grib file ~trx/ecCodes/data/ztuv500.grib



0

3 Obtain the relevant values for the computation out of the u and v grib fields

5 Print (a subset of) the wind speed values computed out of the wind components

7 Produce a new file containing a semantically correct field for wind speed

9 Produce a plot of the new field (using python)

10 Print the 10 points with maximum wind speeds (with their lat/lon coordinates)