Python in Hydro-Meteorology State & Prospects

I learned it last night! Everything is so simple!
Hello world is just print "Hello, world!"

I dunno... dynamic typing? whitespace?
Come join us! Programming is fun again!
It's a whole new world up here!
But how are you flying?

I just typed import antigravity
That's it?
... I also sampled everything in the medicine cabinet for comparison.
But I think this is the Python.
HELLO, WORLD!

SHABANOV PAVEL
research associate
in
P.P. Shirshov Institute of Oceanology RAS, Moscow, Russia

www.ocean.ru
www.sail.msk.ru
www.geofortran.blogspot.ru
PRESENTATION STRUCTURE

PART A
«PYTHON IN EARTH SCIENCES»

PART B
«DATA IN EARTH SCIENCES»

PART C
«MATH IN EARTH SCIENCES: STATE & PROSPECTS»
Python: Programming Language

www.python.org

PYTHON IS A HIGH-LEVEL PROGRAMMING LANGUAGE.

IT SUPPORTS OBJECT-ORIENTED & PROCEDURAL STYLES OF PROGRAMMING
Why Python? PYPL!

The PYPL Index is created by analyzing how often language tutorials are searched on Google.
Why Python?

**ADVANTAGES**

• Free software
• Easy to Learn
• Good I/O
• Cross-platform
• OOP style
• Graphics & maps

**DISADVANTAGES**

• Not very fast
• Python 3 is not compatible with Python 2
• Many scientific modules require Linux OS
• «Big data – big problems»
Scientific «transporter»
Python modules for scientific calculations
http://www.scipy.org/

1. NUMPY
2. SCIPY
3. MATPLOTLIB
4. PANDAS
5. BASEMAP/CARTOPY

-> RESULT <-

Filtering a Delaunay mesh (application to high-resolution tricontouring)
NUMPY: ARRAYS & MATH
PROVIDES ARRAYS & BASIC MATHEMATICAL PROCEDURES WITH THEM (SLICES, RESHAPING, etc.)

MATLAB® and NumPy/SciPy have a lot in common
SCIPY: SCIENTIFIC LIBRARY

PROVIDES SCIENTIFIC Routines & PROCEDURES

- **Clustering** package (scipy.cluster)
- Discrete **Fourier transforms** (scipy.fftpack)
- **Integration** and ODEs (scipy.integrate)
- **Interpolation** (scipy.interpolate)
- Linear **algebra** (scipy.linalg)
- **Optimization** and root finding (scipy.optimize)
- **Signal processing** (scipy.signal)
- **Spatial algorithms** and data structures (scipy.spatial)
- **Statistical** functions (scipy.stats)
PANDAS: TIME SERIES

USING DATAFRAMES TO WORK WITH MASKED ARRAYS & HIGH-DIMENSIONAL DATA, TIME SERIES
MATPLOTLIB: GRAPHICS

PUBLICATION QUALITY GRAPHICS AND PLOTS
BASEMAP/CARTOPY: MAPS

SUPPORTS MAPS & GEOSPATIAL DATA
EXAMPLE 1

CLIMOGRAM FOR MOSCOW DURING XX-XXI
EXAMPLE 2

CHANGES IN ANNUAL PRECIPITATION SUMS OVER RUSSIA FOR THE LATEST 30 YEARS
DATA IN EARTH SCIENCES

IDEALIZED SCENARIO
- INPUT DATA [20%]
- PROCESSING DATA (MATH METHODS) [70%]
- OUTPUT DATA [5-10%]

REALISTIC SCENARIO
- INPUT DATA [70%]
- PROCESSING DATA (MATH METHODS) [25%]
- OUTPUT DATA [5%]

«BIG HEAD PROBLEM»:
A LOT OF TIME & EFFORTS ARE SPENT ON PREPARING DATA
INPUT DATA

READING SCIENTIFIC DATA IN DIFFERENT FORMATS

DATA FORMATS

• **ASCII** (TEXT) – I CAN READ IT & SEE
• **BINARY** – I AM NOT A MACHINE, I CAN NOT READ IT

ANOTHER POINT OF VIEW

• **SELF-DESCRIBING FORMATS** -> netCDF
• **XLS, XLSX** (MS EXCEL)
• **DB** (DATABASES, MySQL)

CLIMATE DATA IS NUMERICAL «BIG DATA».
GIGABYTE FILES «BITE»!
INPUT DATA IN EARTH SCIENCES
QUALITY CONTROL OF RAW EXPERIMENTAL DATA & MEASUREMENTS

- PHYSICAL LIMITS
- FILTER OUTLIERS
- TRANSFER MISTAKES
- MISSING VALUES
- FILLING DATA GAPS
- STORE REVIEWED DATA
INPUT DATA
IN EARTH SCIENCES
INFORMATION ABOUT 4D-DIMENSIONAL EARTH STATE

SPATIAL DATA

• **IRREGULAR** MEASUREMENTS IN POINTS

• **REGULAR (GRIDDED)** OFTEN RECALCULATED FROM IRREGULAR DATA
PROCESSING DATA

RESHAPING, RECALCULATING DATA USING MATH METHODS

INTERPOLATION
TO CONVERT IRREGULAR SPATIAL DATA TO GRID (FOR USING NUMERICAL PROCEDURES) YOU NEED TO INTERPOLATE THE DATA

STATISTICS
TO UNDERSTAND REAL & SIGNIFICANT CHANGES IN PARAMETERS OF EARTH SYSTEM (MEANS, DEVIATIONS & SO ON)
OUTPUT DATA
IN EARTH SCIENCES

STORING BIG DATA ARRAYS IN SPECIFIC FORMATS

• Network Common Data Format (NetCDF)
  http://www.unidata.ucar.edu/packages/netcdf/index.html
  Multi-dimensional arrays, de facto standard for model output for oceans/atmosphere

• Hierarchical Data Format (HDF)
  http://hdf.ncsa.uiuc.edu/
  Parallel I/O, large files
  Collection of file formats
    Scientific Data is like NetCDF
  Used for NASA images
REPRESENTING OUTPUT IN EARTH SCIENCES

- Plots, graphs, tables, but the main are maps

- Interactive plotting
- GIS technologies
- Autoplotting
MATH IN EARTH SCIENCES: STATE

STATISTICAL DISTRIBUTION
1. LINEAR REGRESSION
2. STATISTICAL SIGNIFICANT TESTS

MODELING OF CLIMATE, NWP
1. SOLVING SYSTEM OF LINEAR EQUATIONS
2. INTERPOLATION

SIGNALS IN CLIMATE SYSTEM
1. SIGNAL PROCESSING
2. SPECTRAL ANALYSIS
3. WAVELETS
MATH IN EARTH SCIENCES:
GEOSTATISTICS

KRIGING PROCEDURE
IN RUSSIA DESCRIBED BY GANDIN
AND IS KNOWN AS
«OPTIMAL INTERPOLATION»

http://www.gslib.com/
MATH IN EARTH SCIENCES: PROSPECTS

INTENSIVE WAY

- MATH METHODS HAVE TO BE VALIDATED FOR CURRENT SPECIFIC OBJECTS & AREAS
- STATISTICS NEEDS TO BE USED MORE ACCURATELY
- INTERDISCIPLINARY COLLABORATION

EXTENSIVE WAY

- NEW MATH METHODS USING
- IMPLEMENTING NON STANDART PROCEDURES
- WIDER APPLICATION OF SCIENTIFIC NUMERICAL LIBRARIES
MATH IN EARTH SCIENCES: PROSPECTS

GOOD CHANCE FOR DISCUSSION

THIS SLIDE COULD PRESENT YOUR RESEARCH!
ACTUAL PROBLEMS IN CLIMATOLOGY

OBJECTIVE CLASSIFICATION OF SYNOPTICAL STATES

**NOW:** YOU LOOK AT SYNOPTIC MAPS AND DETERMINE SYNOPTICAL STATE

**FUTURE:** YOU DETERMINE RESEARCH AREA AND GET DATASET OF SYNOPTICAL STATES FROM DATABASE

**INPUT DATA**
SLP GRIDDED DATASETS
REANALYSIS
ECMWF, NCEP/NCAR

**SOLUTION**
IS COMING
NUMERICAL WEATHER PREDICTION

OCCAM’s RAZOR: «OBJECT’s SPATIAL & TEMPORAL SCALES DETERMINE INPUT DATA SCALES»

MAIN PROBLEMS
• LACK OF INITIAL DATA
• MISTAKES IN INITIAL DATA
• MICROPHYSICS
• TECHNOLOGICAL PROBLEMS

WRONG WAYS
• MODEL ST-LAGS SIMPLE INCREASE
• OLD CLUSTERS OR SUPERCOMPUTERS SIMPLE REPLACEMENT BY NEW MORE POWERFUL MACHINES
YET ANOTHER CONFERENCE 2014

MOSCOW, OCTOBER 30
https://events.yandex.ru/lib/talks/2397/

Nicolay Koldunov
Climate Service Center, Hamburg

«Python for interactive data analysis»
http://koldunov.net/
www.oceanographers.ru
www.yandex.ru
Almost all plots & maps in the presentation were implemented using Python.