GEOCOMPUTATION AND MACHINE LEARNING FOR ENVIRONMENTAL APPLICATIONS

**Date: 6th April 2022, 3.00 - 5:45 PM UTC time
26th May 2022, 3.00 – 5:45 PM UTC time**

**Giuseppe Amatulli
Longzhu Shen
Antonio Fonseca**

Learning objectives

With continuous practice through the lectures, students will become familiar with new command lines and cover numerous topics, including:
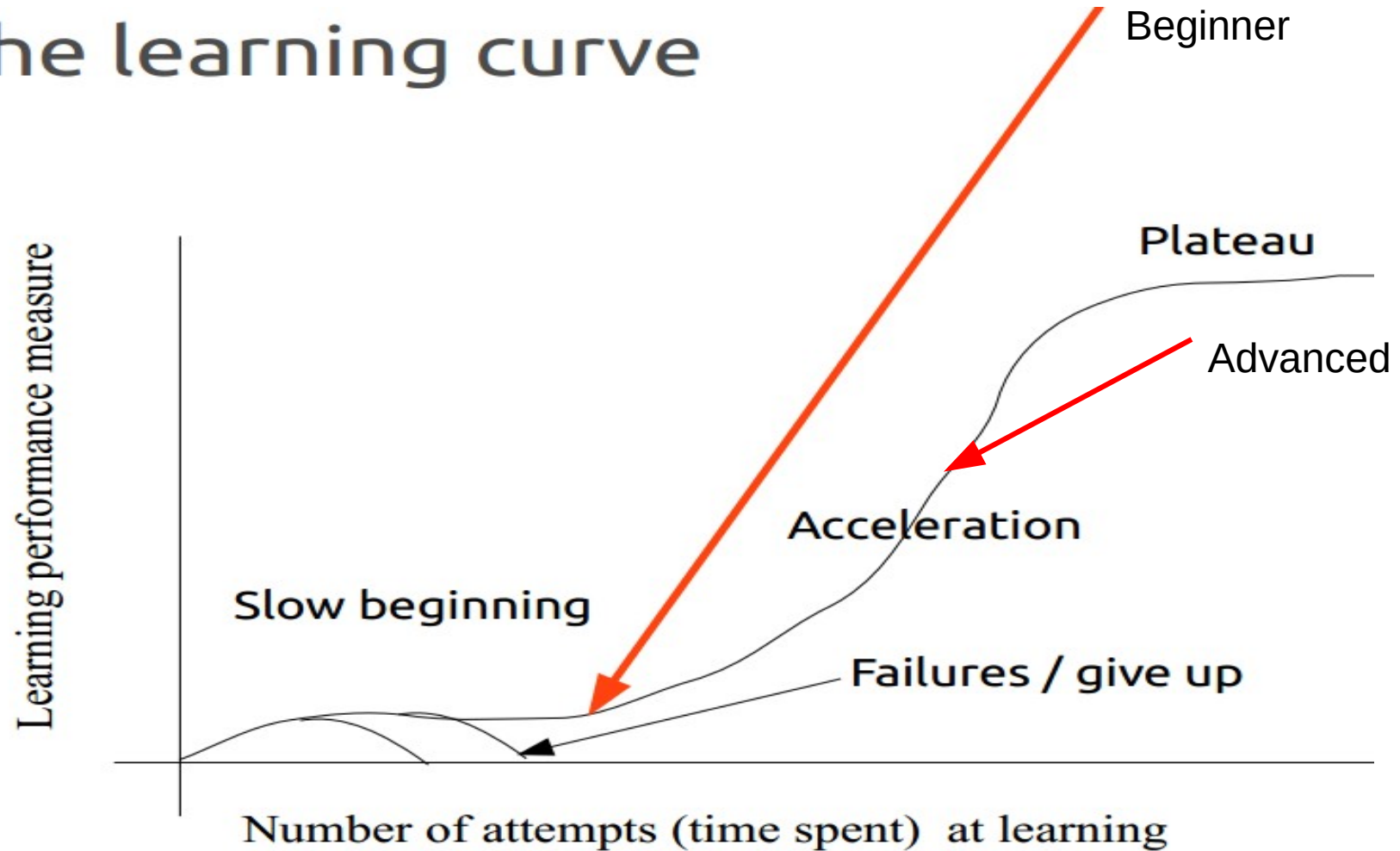
- Learning open source tools for GIS and RS applications.
- Acquiring command line utilities for spatial/temporal data under Linux OS.
- Acquiring command line utilities and ML theoretical foundation for environmental application.
- Developing data processing skills.
- Independent learning, critical thinking, problem solving.

Upon completion of the course, you should be able to:

- Apply the process of science, by conducting, analyzing, and interpreting findings related to GIS & RS project in the framework of ML applications
- Use quantitative reasoning for statistical/spatial analysis
- Convey your understanding of environmental phenomenons

The learning curve

Beginner

Plateau

Advanced

Acceleration

Slow beginning

Failures / give up

Learning performance measure

Number of attempts (time spent) at learning

Scientific knowledge

- **Spatio/temporal analysis**
- **Spatio/temporal data integration**
- **Spatio/temporal modeling**
- **Geostatistic**
- **Machine Learning**

Tools

Grass & Qgis **Geographic Information Systems**

Python: GIS, statistic, modeling, text manipulation

LINUX Bash **shell programming**

AWK **for processing text-based data**

GDAL/OGR/PKTOOLS **geotools library for the manipulation of geospatial data**

Knowing each other (3 min)

- **Name, where are you coming from….**

- **What is your background and personal interest?**

- **Final project / PhD thesis objectives / keywords?**

  - **What data are you going to analyses?**

  - **Not sure yet… no problem**

- **Do you have an experience on Linux OS or other open source software?**

- **Do you currently use any programming language?**

- **What are your interests and expectations on this training?**

Coding knowledge

Covered in the course

- **AWK, GDAL, PKTOOLS, CDO, Python, R, ML (keras, TensorFlow, PyTorch)**

- **Parallel processing in bash and python environment**

- **Machine Learning: theoretical foundation and application in the environmental field**

- **Supervised regression application**

- **Image processing / raster processing / large data-set processing in a proper way**

- **Environmental applications (hydrology, forestry)**
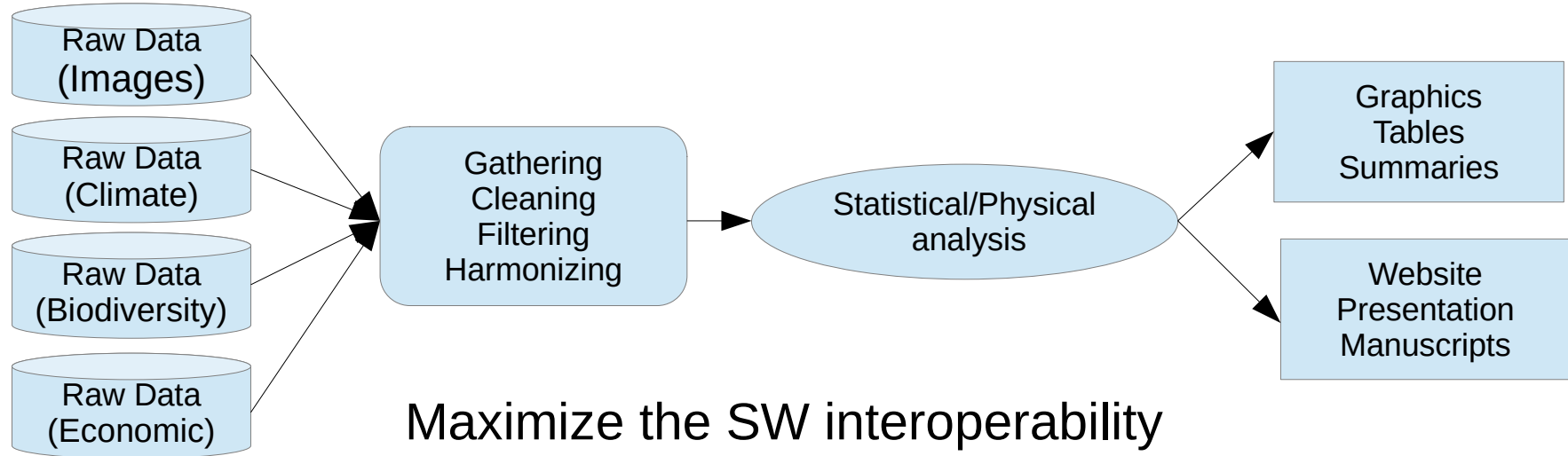
Syllabus clarification

- **Class web-page**
  http://spatial-ecology.net/docs/build/html/COURSESAROUNDTHEWORLD/course_geocomp_ml_04-05_2022.html

- **Data via github** **https://github.com/selvaje/SE_data**

- **Online recording video lecture later shared on**
  http://spatial-ecology.net/docs/build/html/COURSESAROUNDTHEWORLD/course_geocomp_ml_04-05_2022.html

- **Handling script and data via github (code)**

- **Community support among us for trouble shouting via slack**
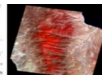
Reproducible research & "big data" processing

Codes that are easily published **> no license constraints**

Complex work-flows **> integrate different data analysis methods**

Raw Data (Images)

Raw Data (Climate)

Raw Data (Biodiversity)

Raw Data (Economic)

Gathering Cleaning Filtering Harmonizing

Statistical/Physical analysis

Graphics Tables Summaries

Website Presentation Manuscripts

Maximize the SW interoperability in a stable Operating System

Why use Linux/OpenSource?

Security: extremely stable and reliable, no viruses, interoperable: Unix, Windows, Mac, Android, …

Applications: thousands of free programs, programming languages, server services

Versatility: minimum HW requirements, extremely portable, very fast performance

Freedom: free to download/test/install/modify/ configure/develop/distribute/... it's fun!

Freedom? and why it's fun?

Code:
- Understating the code beyond a process
- Be able to modify the code
- Build up your own algorithm.
- Use all the SW that I want without license constraints

Help:
- Get help from mailing list
- Keep in touch with the developers for code adjustment and improvement

Process:
- Job priority processing
- Job scheduling
- RAM management

Remote server:
- Automatic connection to remote servers
- Overpassing quota issues in remote servers, by creating a folder linked to your PC
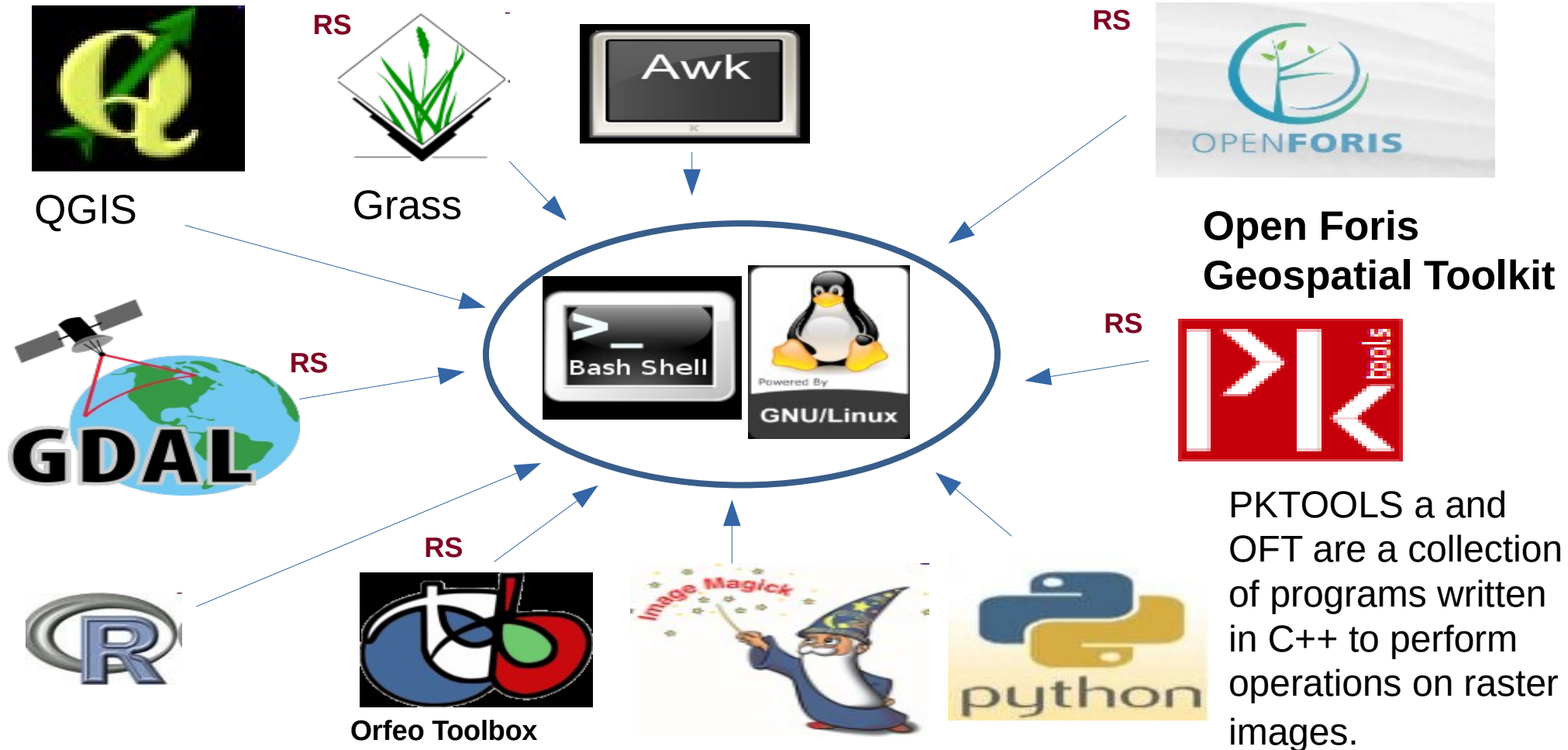
Hardware resources:
- Storing temporal file in ram rather in the hard-disk, by creating a folder in the ram
- Get the best of different programing languages and create a unique work flow.

Last but not least:
- Enjoy the life in the meantime the PC is working for you!

Ubuntu Linux operating system

Programming languages interaction

QGIS

Grass

RS

RS

Open Foris Geospatial Toolkit

RS

RS

PKTOOLS a and OFT are a collection of programs written in C++ to perform operations on raster images.

RS

Orfeo Toolbox

Ubuntu Linux operating system
Programming languages interaction

QGIS

Grass

Open Foris Geospatial Toolkit

PKTOOLS a and OFT are a collection of programs written in C++ to perform operations on raster images.

Orfeo Toolbox

Ubuntu Linux operating system
Programming languages interaction

QGIS

Grass

Open Foris Geospatial Toolkit

PKTOOLS a and OFT are a collection of programs written in C++ to perform operations on raster images.

Orfeo Toolbox