Training Program for ML-PREP

MARK A. ASCH, Université de Picardie Jules Verne, FRANCE

This document presents all the details of the training sessions to be held in the framework of the ML-PREP 2025-26 research program. The training will cover geostatistics, geospatial data analysis, machine learning and seismic wave propagation. The objective is to use the training as a basis for addressing pertinent research questions for seismically induced landslides.

Additional Key Words and Phrases: Geostatistics, Machine learning, Wave propagation

OVERVIEW AND PRE-REQUISITES

This training course for the ML-PREP team, is planned over 3 days, combining

- theoretical lectures,
- practical applications and coding exercises,
- discussion/formulation of research projects.

The pre-requisites for the training are:

- basic machine learning as presented in [1];
- elements of geographic data processing and geostatistics—see [2];
- basic numerical analysis, including matrix linear algebra, numerical integration, solution of ODEs—see [6];
- some familiarity with python, R and unix (command-line).

The Table below presents the global structure and provides clickable links to all the training material located in the github.

Day	Topic	Links
1	Machine Learning	Lectures, Examples
2	Geostatistics	Lectures, Examples
3	Wave propagation	Lectures, Examples
4	Research projects	TBC

1 DAY 1: MACHINE LEARNING

Objective: we will address practical aspects of the use and implementation of machine learning.

1.1 Pre-requisites

- Basics of Machine Learning: lecture notes and examples are here.
- A python environment—see below.

1.2 Software environment

The following commands will install the good coding environment for machine learning.

```
conda create -n prep_ml python=3
conda activate prep_ml
conda install jupyterlab numpy matplotlib pandas
conda install scikit-learn
conda install pytorch torchvision -c pytorch
pip install islp
.
```

This work is supported by DOST-PCIEERD, Philippines.
Author's address: Mark A. Asch, Université de Picardie Jules Verne, 33 rue Saint Leu, Amiens, 80039, FRANCE.

```
.
jupyter notebook
.
.
conda deactivate
```

Some remarks:

- Full software setup instructions are available in the setup document.
- The name of the environment prep_ml can be modified, as you like.
- The use of `pytorch` is optional, and not indispensable. There are 2 tutorials available:
  - basics of tensors - pytorch_101
  - machine learning - pytorch_102
  - use of a gpu - pytorch_M2

1.3 Lectures—morning session Day 1.

An introductory lecture, "AI/ML for Science, Science for AI/ML" will set the scene and describe the overall stakes, in terms of

- Science.
- Society.
- The future of knowledge.

The advanced lecture notes themselves cover:

- how to choose a machine learning method?
- cross-validation and tuning of machine learning models
- evaluation and performance metrics
- causality and correlation
- features and model selection
- PINN (physics inspired neural networks)

1.4 Examples, Use-cases and Exercises—afternoon session Day 1.

The examples here, cover two practical aspects of the use of ML. Many more examples of ML will be encountered in the training of Day 2 (geospatial data analysis) .

Topic	Links
Nested cross-validation combining tuning and performance evaluation.	Notebook
Finalizing, saving and reusing a cross-validated model.	Notebook
Pytorch intro to tensors.	pytorch_101
Pytorch intro to machine learning.	pytorch_102
Pytorch intro to gpu's	pytorch_M2

2 DAY 2: GEOSTATISTICS AND GEOSPATIAL DATA ANALYSIS

2.1 Pre-requisites

- Basic course (github directory) - see the table below in Section 2.4 for individual, detailed links.
- Python environment based on `geopandas`—see below.
- Rstudio and `mlr3` environment

2.2 Software environment

We require a `conda` environment, with a number of packages.

```
conda create -n geo_env
conda activate geo_env

conda config --env --add channels conda-forge
conda config --env --set channel_priority strict

conda install python=3 geopandas geodatasets
conda install plotly rasterio osmnx contextily
conda install jupyter
```

For the use-case on landslides, we need an R environment with

- R
- Rstudio
- mlr3 package

2.3 Lectures—morning session Day 2.

Topic	Links
Geostatistics • probability and stochastic processes • variograms • kriging	lecture notes
Geospatial data analysis and machine learning • model evaluation • spatial autocorrelation • spatial cross-validation • use-cases	lecture notes

2.4 Examples and Exercises—afternoon session Day 2.

Topic - basics	**Links**
Data - loading vector data with geopandas.	Notebook 0
Geometries - use shapely library for creating, manipulating, and analyzing geometric objects.	Notebook 1
Vector Data -load vector data, use the Python library geopandas, an extension of the popular data manipulation library pandas, combined with shapely's geometry processing capabilities.	Notebook 2
Visualizing Geospatial Data - use matplotlib together with geopandas to create detailed, engaging, and insightful geospatial visualizations that can be applied to a wide range of data analysis tasks	Notebook 3
Map Projections - overview of map projections and how to effectively manage CRS (coordinate reference systems) in geospatial projects, ensuring that spatial analyses and visualizations are both accurate and meaningful.	Notebook 4
Raster Data - explore techniques and tools in Python to handle large raster datasets. Clipped data file to be downloaded: https://tinyurl.com/cp4ey9cn	Notebook 5
Introduction to OpenStreetMap Data - download and visualize different types of OpenStreetMap data.	Notebook 6

Topics - advanced	**Links**
Geostatistics: stochastic processes, stationarity, variograms, simple kriging	notebook
Kriging using `pykrige`	Notebook
Kriging with CV using `pykrige`	Notebook
Kriging + ML (exercise)	Notebook
Raster analysis of slopes and slope stability using `rioxarray`	Notebook
Raster analysis of slopes and slope stability generated by LLM, using only `rasterio`	Notebook
Collection of more detailed raster operations from AutoGIS (optional)	Notebook
Spatial resampling (CV) with k-fold clustering on housing data - data file	Notebook
Use-case 1: Ecuador landslide susceptibility analyis (in R) using spatial CV and spatial tuning	directory, Rmd file
Use-case 2: Domestic violence with spatial CV and random forest	notebook, data, paper
Use-case 3: Obesity prevalence with spatial CV and MLP neural network	notebook, data, paper

3 DAY 3: WAVE PROPAGATION

3.1 Pre-requisites

- Basic numerical analysis—see [5], [6] and references therein.
- Conda/Python environment—see below.
- SPECFEM2D—see below.

3.2 Software environment

3.2.1 Python. A very basic python environment is needed for the first part:

```
conda create -n prep-wave
conda activate prep-wave

conda install jupyterlab numpy scipy matplotlib
jupyter lab
.
.
.
conda deactivate
```

3.2.2 SPECFEM. The spectral finite element code that we will use is SPECFEM2D. Note that to run the SPECFEM package on Windows rather than on Unix machines, you can install Docker or VirtualBox (installing a Linux in VirtualBox in the latter case) and run it easily from inside them.

To install and run SPECFEM under unix (or macos), do the following.

- Download the software:

```
git clone --recursive --branch devel https://github.com/SPECFEM/specfem2d.git
```

- Go into the specfem2d directory, configure the Fortran and C compilers and compile the code:

```
./configure FC=gfortran CC=gcc
make all
```

- For a quick test, run the default example with these commands:

```
./bin/xmeshfem2D
./bin/xspecfem2D
```

and check the output files in ./OUTPUT_FILES/

3.3 Lectures—morning session Day 3.

The theoretical lecture notes cover the following subjects:

- basics of seismic wave propagation: harmonic waves, acoustic waves, seismic waves
- finite difference method
- finite element method
- spectral element method

3.4 Examples, Use-cases and Exercises—afternoon session Day 3.

Topic	Links
1D vibration ODE	notebook
2D acoustic wave equation in heterogeneous medium.	notebook1 notebook2
Animation tutorial.	notebook
1D spectral element code in a homogeneous medium.	notebook
1D spectral element code in a heterogeneous medium.	notebook
SPECFEM examples (see above for initial test):	
Calculate the spectrogram of a sinusoidal function.	notebook
Display seismograms from SPECFEM.	notebook
Specfem2d: wavefield of 4-layer medium, curved interfaces	directory
Specfem2d: homogeneous half-space	directory
Specfem2d: homogeneous half-space, parameter modification	directory
Specfem3D: homogeneous half-space	inside Docker...
Specfem3D: 3-layer half-space	inside Docker...

4 DAY 3-4: RESEARCH PROJECTS

The topics for research projects will be discussed with PHIVOLCS. There are 3 tentative propositions detailed below. Note that each project relies on the contents of the training program detailed above, and the lectures, examples will need to be restudied and completed as the research progresses. The work on the projects will be divided into teams, with each team concentrating on one of them. Cross collaboration between the teams is strongly encouraged. All projects will be guided by senior staff from CSU and PHIVOLCS, for the duration of the project.

4.1 Project 1: EIL Inventory using ML

A preliminary formulation for this project can be consulted here and here (training version).

- Use tree-based ML on a landslide inventory to generate a detailed susceptibility map.
- Dashboards (interactive) for DRRM and risk scenario exploration.
- Use ML to update the hazard maps by integrating real-time data and continuously (re)learning from new events.
- Post Qualification (Landslide Inventory): (PE)ML can be used to validate and update the hazard maps by comparing predicted hazards with actual landslide occurrences.

4.2 Project 2: ML for Soil Parameters and FoS

A preliminary formulation for this project can be consulted here.

- Use PEML to predict the FoS by taking into account the Newmark physics formulas.
- Perform a feature importance study on the 7 covariates (geotechnical parameters in Newmark).
- Perform a causality study, based on Shapley values, and integrate the results into the hazard/susceptibility map.
- Evaluate predictive performance of past forecasts and hazard maps.

4.3 Project 3: SEM-Newmark Coupling for EILs

A preliminary formulation for this project can be consulted here.

- Step 1: Simulate seismic wave propagation using SEM.
- Step 2: Predict landslide using Newmark displacement analysis.
- Step 3: Couple the two above steps to predict Earthquake Induced Landslides. Study effects of duration of seismic events as well as peak amplitude. Evaluate risks of extreme events.

REFERENCES

[1] M. Asch. CSU-IMU-2023 basic course on machine learning. https://sites.google.com/view/csu2023/basic-course
[2] M. Asch. Elements of geographic data processing in python. https://tinyurl.com/32wsnyz6
[3] G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor. *An Introduction to Statistical Learning*. Springer. 2023. ISL Site
[4] C. Wikle, A. Zammit-Mangion, N. Cressie. *Spatio-Temporal Statistics with R.* CRC Press. 2019. https://spacetimewithr.org/
[5] H.P. Langtangen. *Finite Difference Computing with PDEs: A Modern Software Approach.* Springer Cham. 2017. Download pdf
[6] M. Asch. *A Toolbox for Digital Twins: From Model-Based to Data-Driven.* SIAM. 2022. (available in CSU Library)
[7] H. Igel. *Computational Seismology: A Practical Introduction.* Oxford University Press, 2017. website

ACKNOWLEDGMENTS

The authors would like to thank PhiVOLCS, Caraga State University. This work is supported by the DOST-PCIEERD under Grant No.: 1234567.