

# HERRAMIENTAS DE MINERÍA DE DATOS PARA CIENCIAS DE LA TIERRA

Vorobioff J.<sup>ab</sup>, Pintos T.<sup>a</sup>, Checozzi F.<sup>ab</sup>, Morgada M.<sup>a</sup>, Boggio N.<sup>abcd</sup>, Rinaldi J.<sup>d</sup>, Filipussi D.<sup>ab</sup>, Ristori P.<sup>bc</sup>, Rinaldi C.<sup>abcd</sup>

<sup>a</sup>Departamento de Micro y Nanotecnología, Centro Atómico Constituyentes, CNEA, Argentina

<sup>b</sup>Facultad Regional Buenos Aires, Universidad Tecnológica Nacional, Argentina

<sup>c</sup>CONICET, Argentina, <sup>d</sup>Universidad Nacional de San Martín, Argentina

e-mail: [vorobioff@cnea.gov.ar](mailto:vorobioff@cnea.gov.ar)

## Resumen

- -Minería de datos aplicado a las ciencias de la tierra: no siempre se utilizan las herramientas más adecuadas.
- Muchos algoritmos se utilizan en un único laboratorio y otros grupos de trabajo no lo pueden utilizar.
- Experiencia: LIBS aplicado a suelos, olfatometría, mediciones ambientales e imágenes satelitales.
- Mejorar la velocidad de procesamiento y bajar la carga computacional, minimizar el error, tener algoritmos robustos y buena generalización ante variaciones y ruidos de las entradas.
- Algoritmos flexibles y de fácil manejo para los usuarios, con software de licencia libre e interfaces de usuario amigables, por ejemplo, poseer una interfaz de usuario gráfica (GUI). Las herramientas de software deben contener diferentes métodos de procesamiento, preprocesamiento, validación cruzada y comparación de resultados entre diferentes técnicas.
- Visualización de los datos y los resultados debe ser adecuada.
- Se utiliza inteligencia artificial y métodos estadísticos explicados en el libro del autor.
- Matlab resulta potente, flexible, pero costoso y muchas veces lento. Python resulta flexible y con licencias libres, multiplataforma, sin embargo, los programas no siempre resultan fáciles de usar y modificar para usuarios no programadores, recomendamos utilizar GUI.
- Librerías TensorFlow, PyTorch, Keras y Scikit-learn. Las herramientas RapidMiner y Orange trabajan con diagramas de bloques. Bloques con código propio en lenguaje Python.
- Para que los programas perduren hay que tener en cuenta todos los factores mencionados a la hora de elegir la herramienta más adecuada.

## Resultados

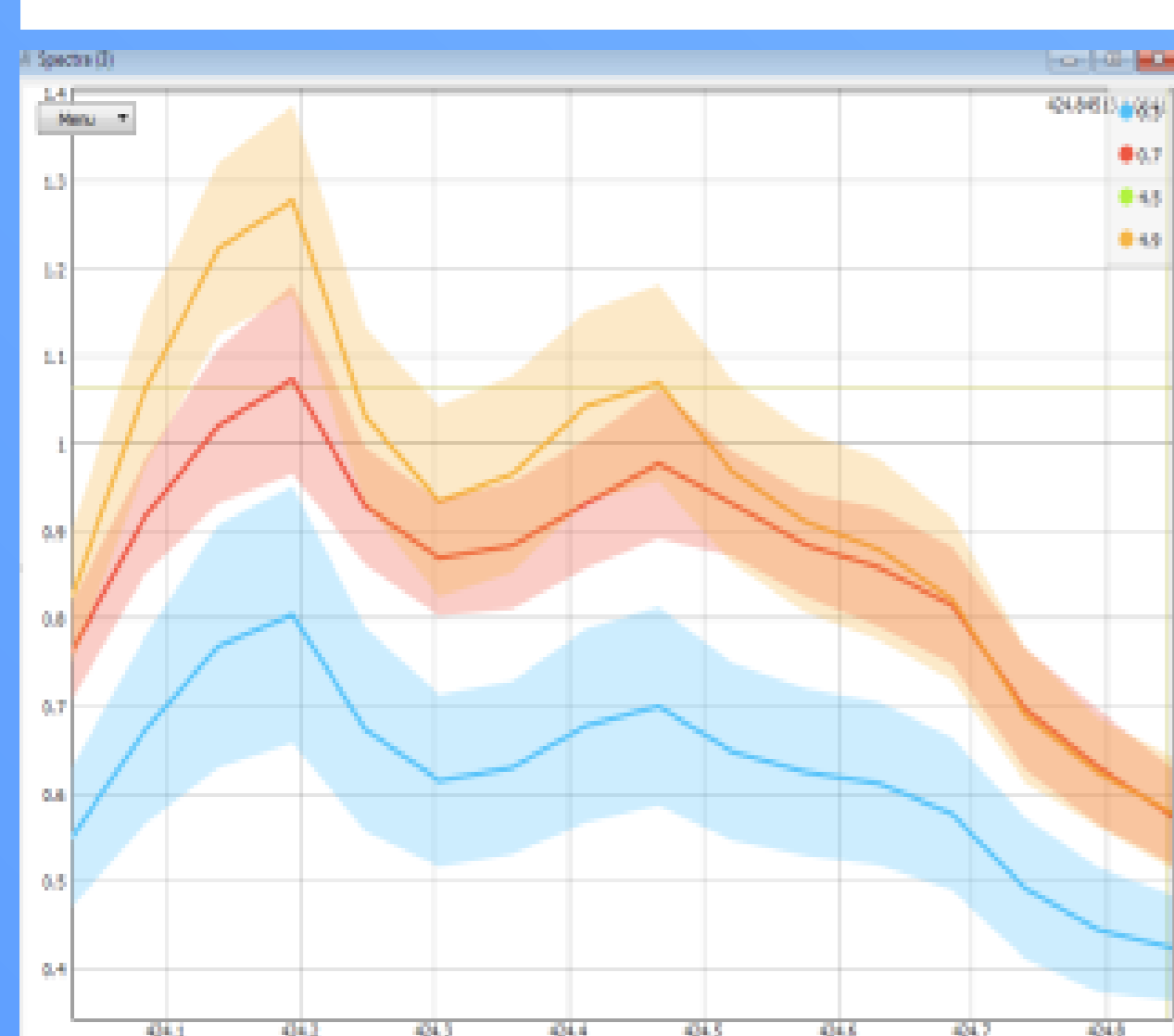
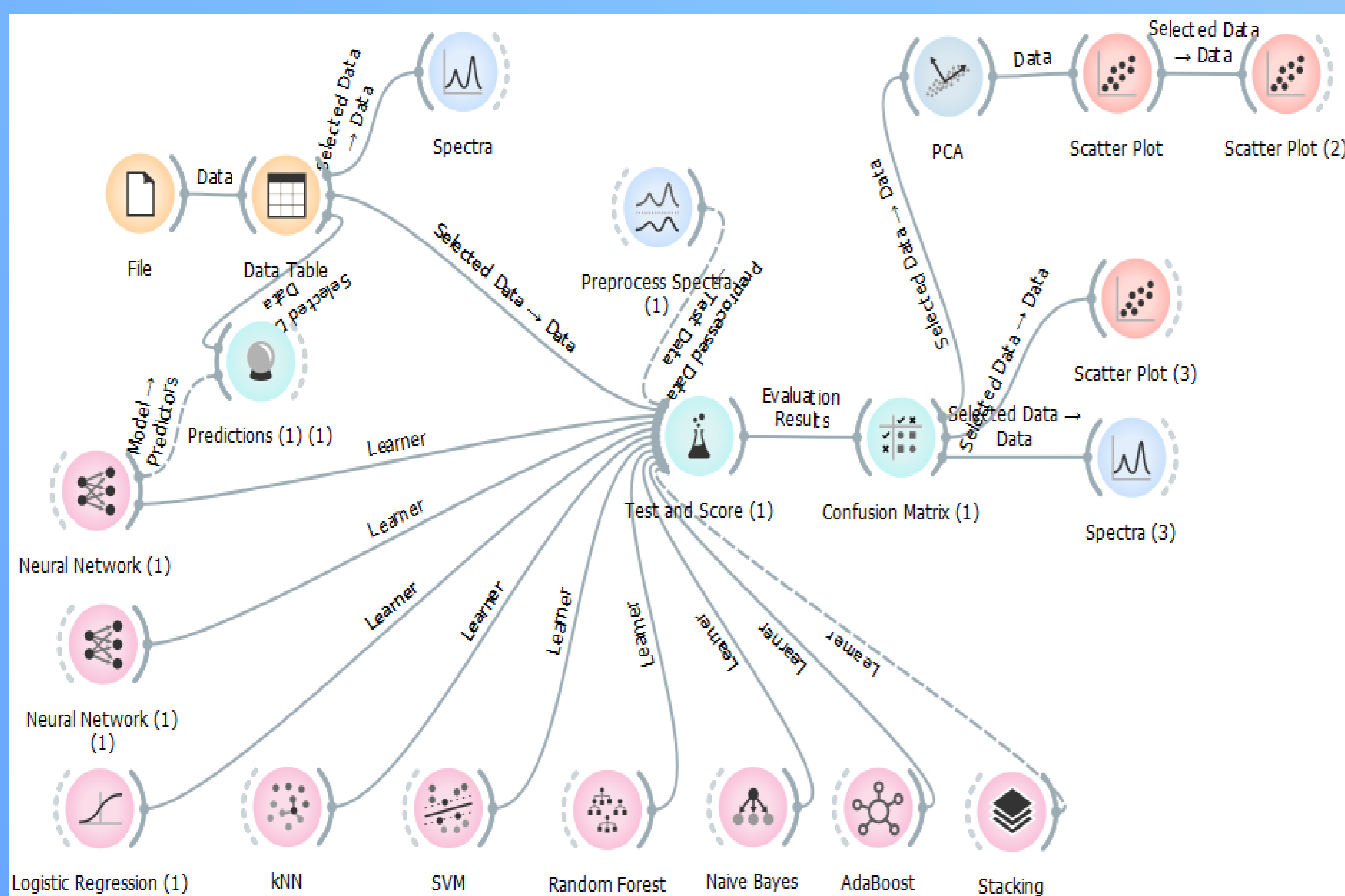


Fig. 3 – Promedio de 59 espectros, 4 grupos

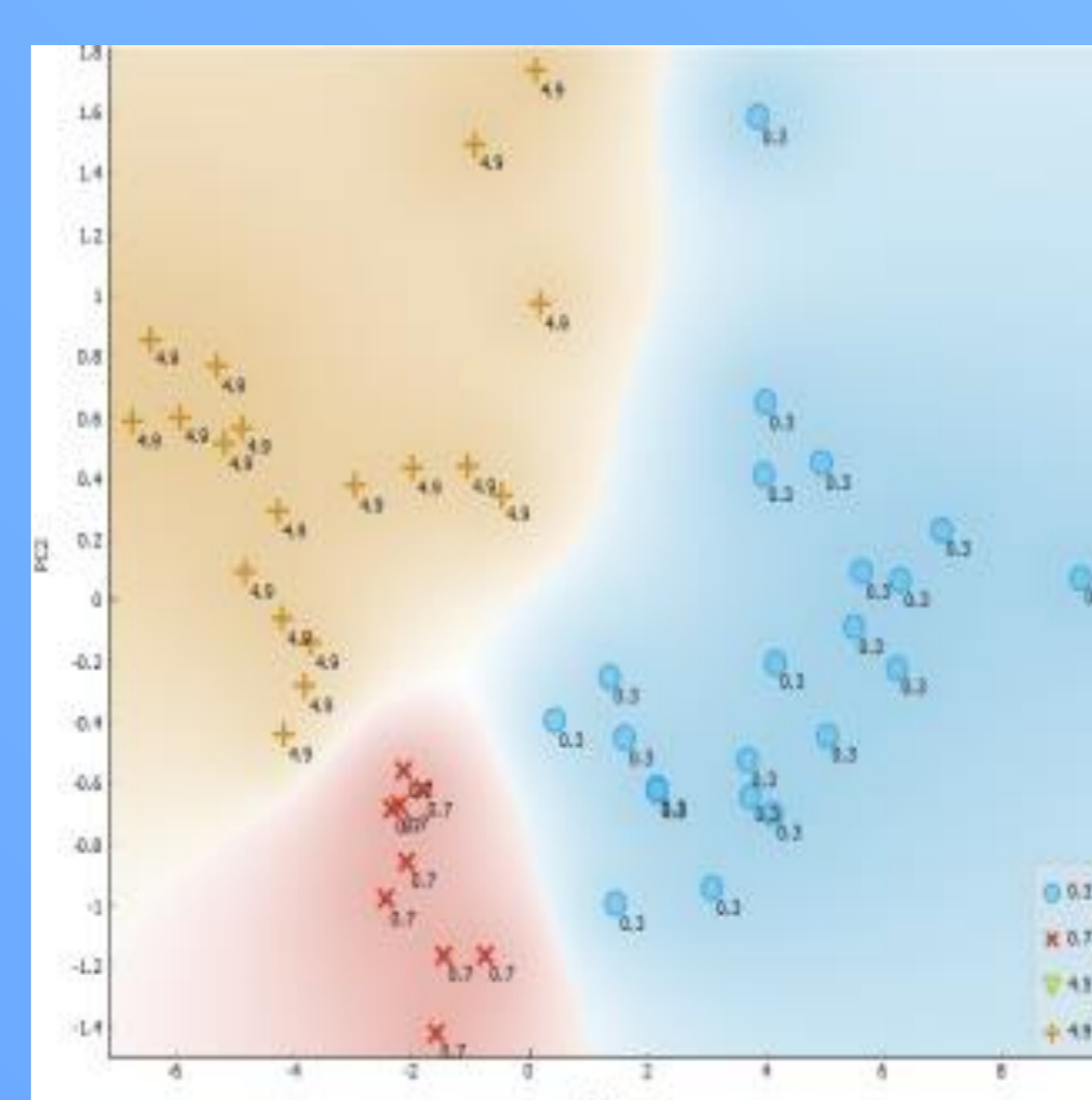


Fig.2 – Resultados PCA

Fig.1 – Izquierda - Ejemplo de procesamiento de datos con Orange



Fig.4 – Software de Reconocimiento de Patrones

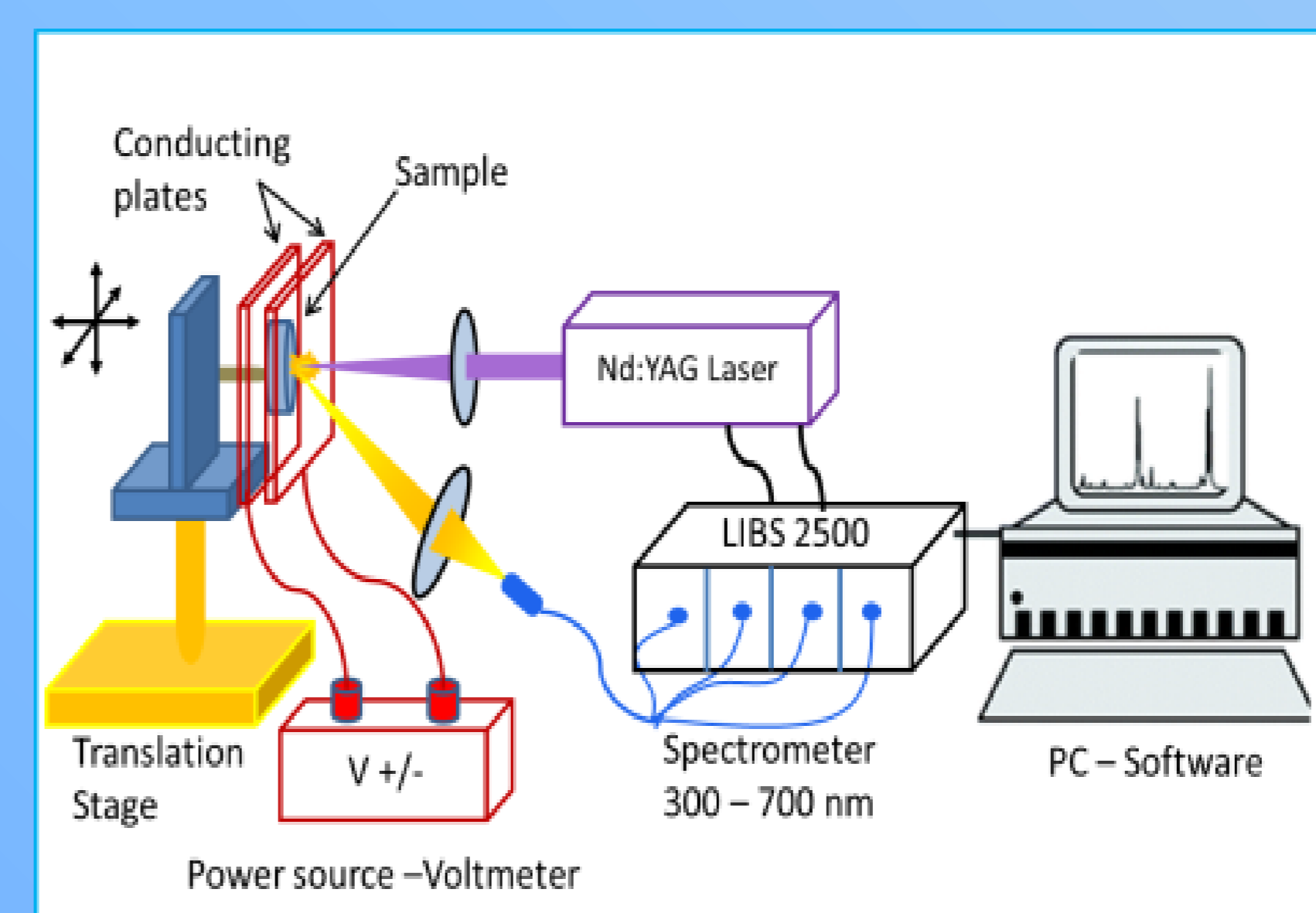


Fig.5 – Experimental Setup

### Referencias

- [1] Phillips, J. D. & Marion, D. A.. Geoderma 141, 89–97 (2007).  
[2] Yu, K.-Q. et al.. Sci. Rep. 6, 27574; doi: 10.1038/srep27574 (2016).