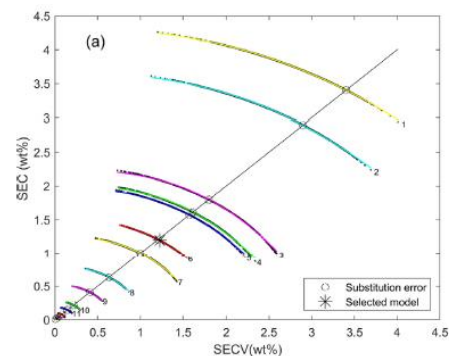


## UNA ESTRATEGIA PARA SELECCIONAR UNA PARTICIÓN DE DATOS DE CALIBRACIÓN EN UN MODELO PLSR

La regresión por mínimos cuadrados parciales (*PLSR*, *Partial Least Squares Regression*) es una herramienta notable en dominios que demandan modelos numéricos de predicción y clasificación de datos, entre otros, quimiometría y econometría. Como toda regresión, *PLSR* requiere datos para la calibración y la validación de su desempeño, y su éxito radica, como siempre, en la calidad y cantidad de datos que se manejen, de manera que se asegure un adecuado número de grados de libertad, así como aplicar con rigor las métricas para evaluar el desempeño.

En esta exposición se presenta una estrategia, apoyada en la correlación entre el error estándar de calibración y el error estándar de validación cruzada, para seleccionar una partición de datos (de calibración) que evita el sobreajuste (*overfitting*) y mantiene valores razonables en las figuras de desempeño del modelo.



**Fecha de sustentación:** febrero 29 de 2024.

### Actividad

Un grupo de estudiante podría proponer algún tipo de modelo predictivo y verificar lo expuesto en la exposición.

### Lecturas de profundización:

A los estudiantes interesados, con todo gusto, se les dará copia de artículos publicados en los que la estrategia se ha implementado.

### Colaboradores:

Este tópico es el resultado de discusiones y trabajo conjunto con los profesores:

Rafael Cabanzo Hernández (Escuela de Física)

Enrique Mejía Ospino (Escuela de Química)