



Food and Agriculture  
Organization of the  
United Nations

**GLOSOLAN**  
Soil spectroscopy  
training workshops

# Aspectos básicos del análisis de suelos mediante espectroscopía de infrarrojo cercano

César Guerrero  
Universidad Miguel Hernández (UMH)  
29 Mayo, 2024

Online  
webinars



# Aspectos básicos del análisis de suelos mediante espectroscopía de infrarrojo cercano

**César Guerrero Maestre**

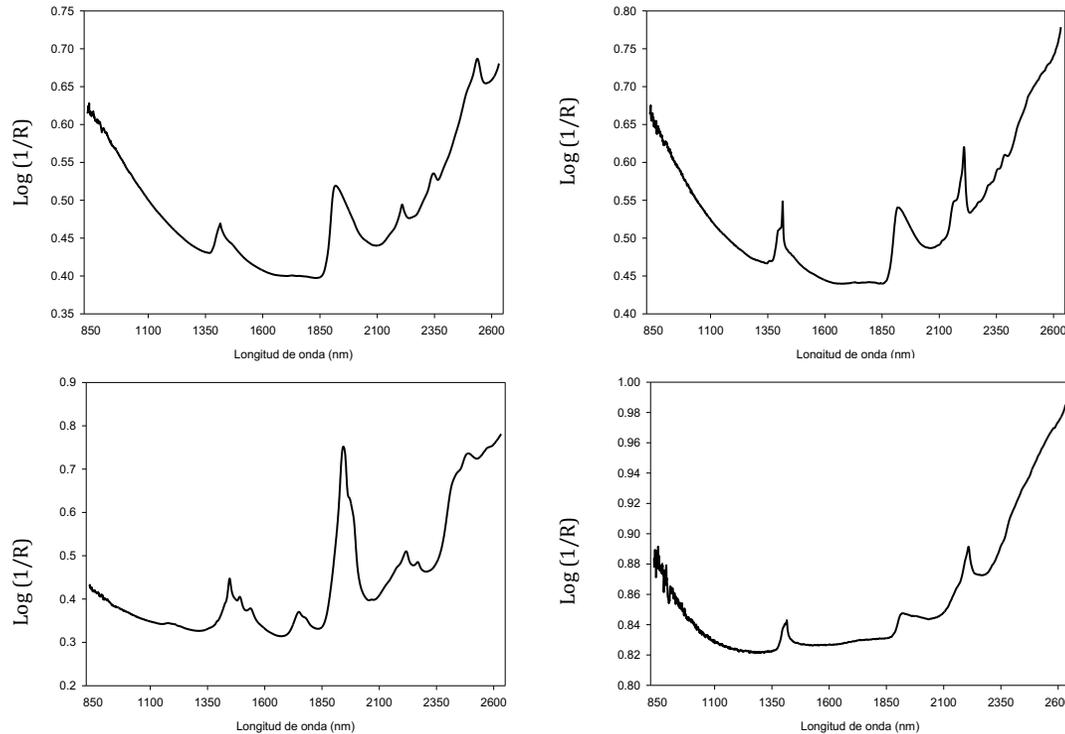
Departamento de Agroquímica y Medio Ambiente  
Universidad Miguel Hernández de Elche ([www.umh.es](http://www.umh.es))

Avda. de la Universidad, s/n. Elche (Alicante) E-03202 . España

Teléf.: +34 966658333. E-mail: [cesar.guerrero@umh.es](mailto:cesar.guerrero@umh.es)

ORCID: <https://orcid.org/0000-0002-3960-9724> 

## Espectros NIR (*Near InfraRed*)



- Suelos {
- Propiedades físicas
  - Propiedades químicas
  - Propiedades biológicas



### Espectroscopía NIR:

- Rápida (<1 minuto)
- Sencilla (simple)
- Limpia (sin reactivos -> sin residuos)
- Barata
- Precisa (repetitivo)
- No destructiva
- Múltiples análisis
- *Empírico (modelos <- librerías espectros)*

## Modelos univariantes

$$y = f(x)$$

*variable dependiente = f (variable independiente)*

$$\text{Peso árbol (y)} = \text{coef.} \cdot \text{Altura árbol (x)}$$

$$\text{Propiedad edáfica (y)} = \text{coef.} \cdot \text{NIR (x)}$$

$$\% \text{ Carbonatos} = \text{coef.} \cdot \text{Absorbancia } (\lambda_{2355 \text{ nm}})$$

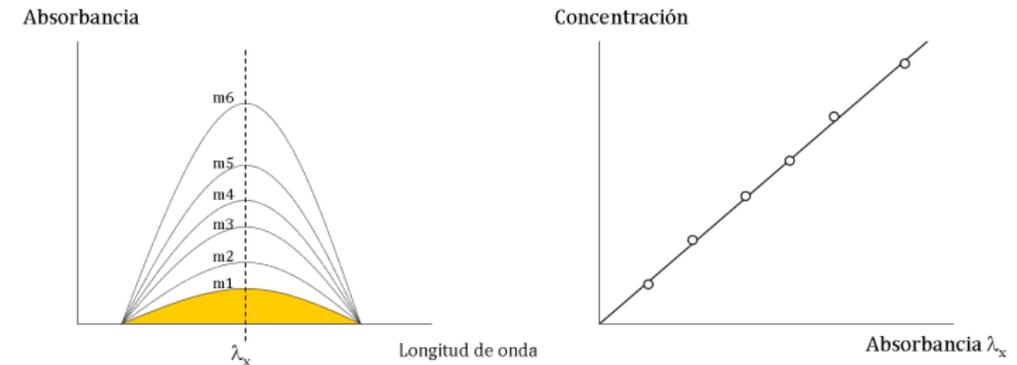
## Modelos multivariantes

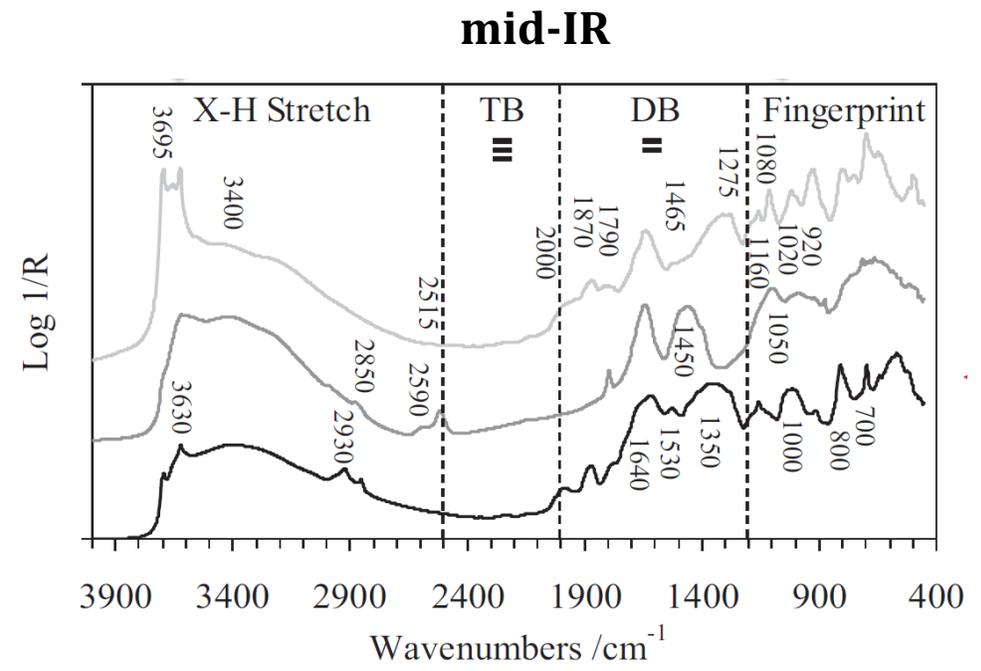
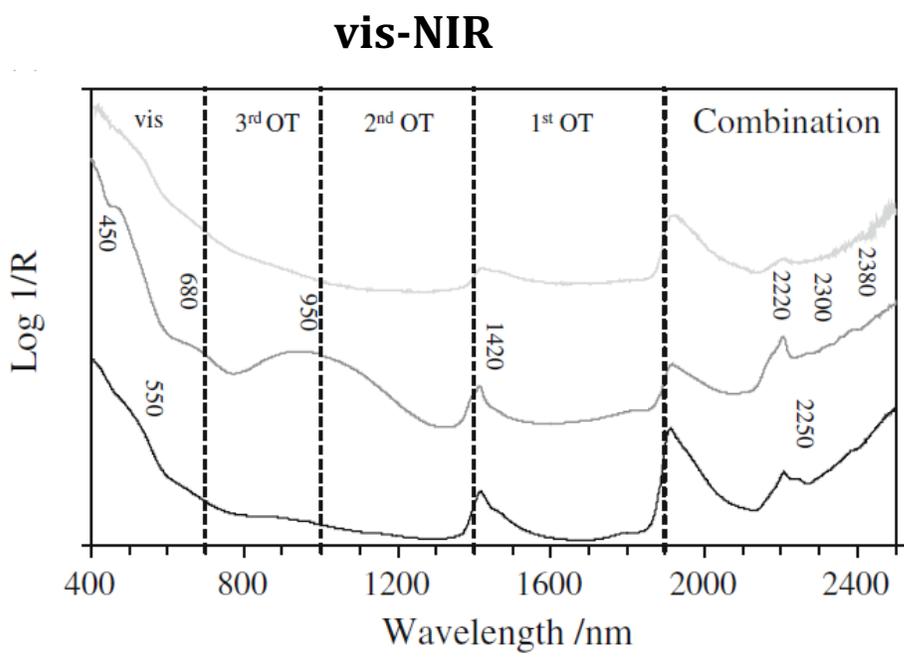
$$Y = \text{coef}_1 \cdot X_1 + \text{coef}_2 \cdot X_2 + \text{coef}_n \cdot X_n$$

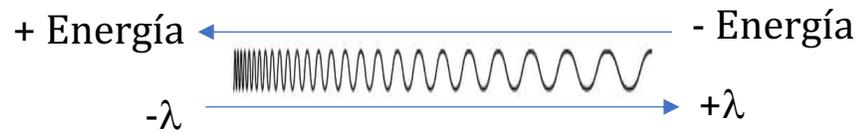
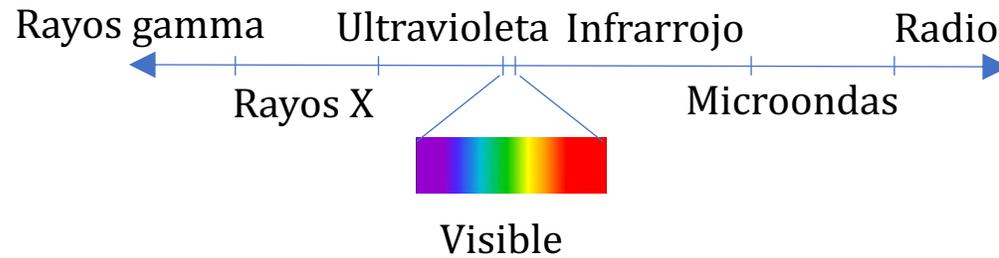
$$\text{Peso árbol} = \text{coef}_1 \cdot \text{Altura árbol} + \text{coef}_2 \cdot \varnothing \text{ tronco}$$

$$\text{Peso árbol} = \text{coef}_1 \cdot \text{Altura árbol} + \text{coef}_2 \cdot \varnothing \text{ tronco} + \text{coef}_3 \cdot \# \text{ ramas} \dots$$

$$\% \text{ Carbonatos} = \text{coef.} \cdot \text{Absorbancia } (\lambda_{2355 \text{ nm}}) + \text{coef.} \cdot \text{Absorbancia } (\lambda_{2535 \text{ nm}}) + \dots$$







$$c = \lambda \cdot \nu$$

$c$  : velocidad de la luz (299792458 m/s)  
 $\nu$  : frecuencia  
 $\lambda$  : longitud de onda

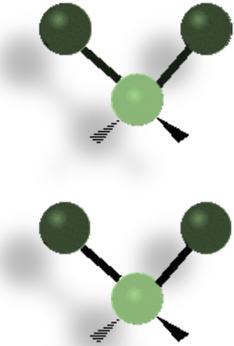
Energía (E)  
 $h$  : cte. Planck  
 $E = h \cdot \nu = h \cdot c / \lambda$

Número de onda	26315 cm <sup>-1</sup>	12820 cm <sup>-1</sup>	4000 cm <sup>-1</sup>	250 cm <sup>-1</sup>	10 cm <sup>-1</sup>
	visible	Infrarrojo cercano	Infrarrojo medio	Infrarrojo lejano	
	Transiciones electrónicas	Sobretonos ( <i>overtones</i> ) y combinación de bandas	Vibraciones fundamentales	Rotaciones moleculares	
	VIS	NIR ( <i>Near InfraRed</i> )	Mid-IR	FIR ( <i>far IR</i> )	
Longitud de onda	380 nm	780 nm	2500 nm	40000 nm	1·10 <sup>6</sup> nm

← NIR ← | → SWIR →  
 ~1100 nm

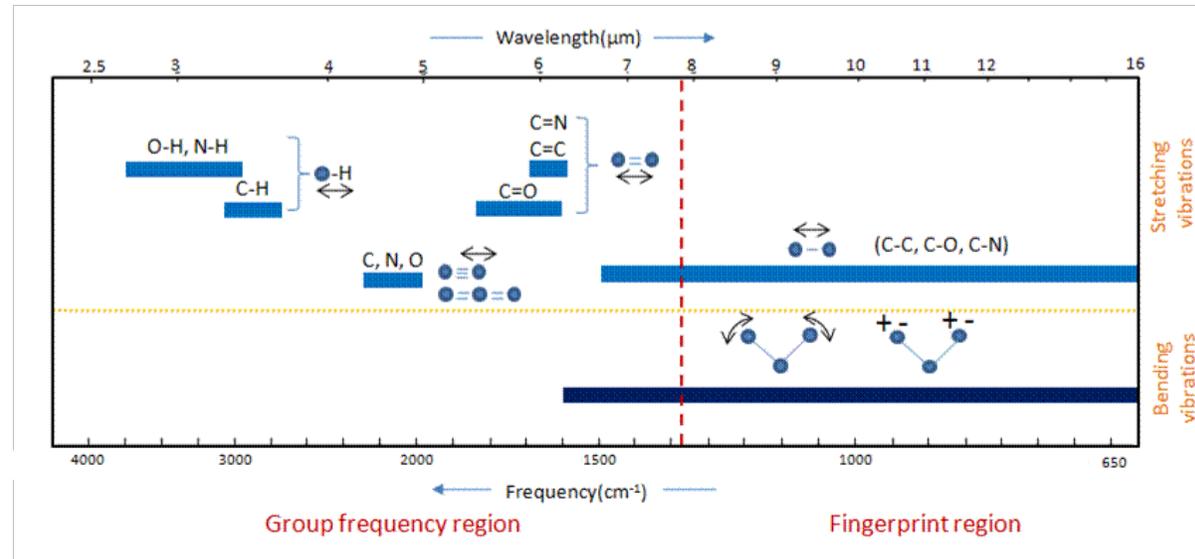
$\nu$  (cm<sup>-1</sup>) = 10<sup>7</sup> /  $\lambda$  (nm)  
 $\lambda$  (nm) = 10<sup>7</sup> /  $\nu$  (cm<sup>-1</sup>)

Estiramiento o tensión  
( $\nu$  stretching)

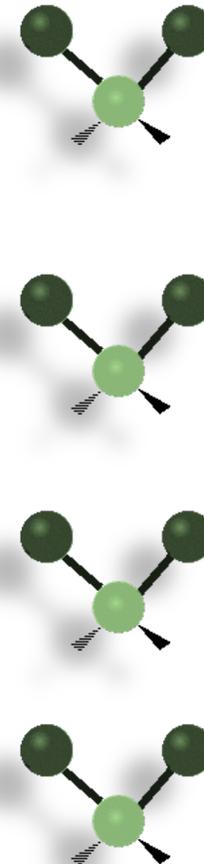


$$\tilde{\nu} = \frac{1}{2\pi c} \sqrt{\frac{K_{x-y}}{\mu}}$$

← Fuerza enlace  
← Masa reducida



Flexión  
( $\delta$  bending)



1691  $\text{cm}^{-1}$  (**vibración fundamental**) Mid-IR  
3530  $\text{cm}^{-1}$  (**vibración fundamental**) Mid-IR

- **Combinación** de bandas:

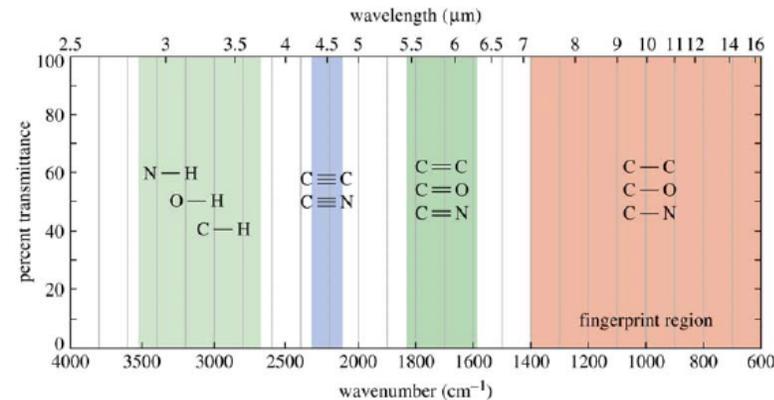
1691 + 3530 = 5221  $\text{cm}^{-1}$  (~1915 nm, NIR)

- **Primer sobretono** (*1st overtone*):

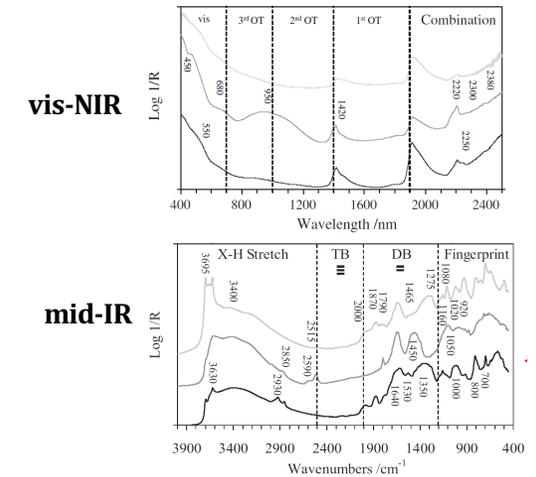
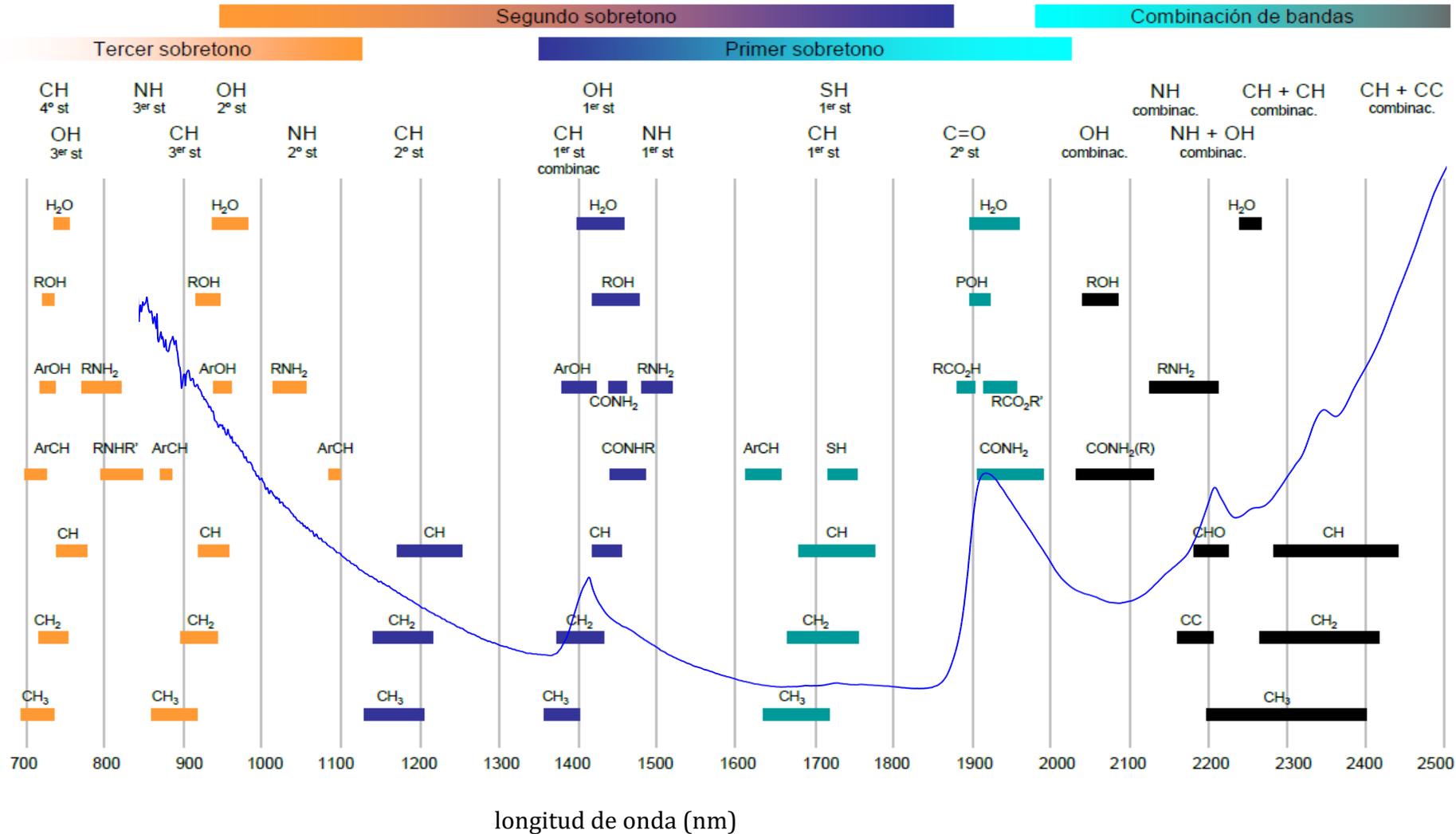
2\*(3530)=7060  $\text{cm}^{-1}$  (~1416 nm, NIR)

- **Segundo sobretono** (*2nd overtone*):

3\*(3530)=10590  $\text{cm}^{-1}$  (~944 nm, NIR)



# Combinación de bandas y sobretonos (NIR)



1691 cm<sup>-1</sup> (**vibración fundamental**) Mid-IR  
 3530 cm<sup>-1</sup> (**vibración fundamental**) Mid-IR



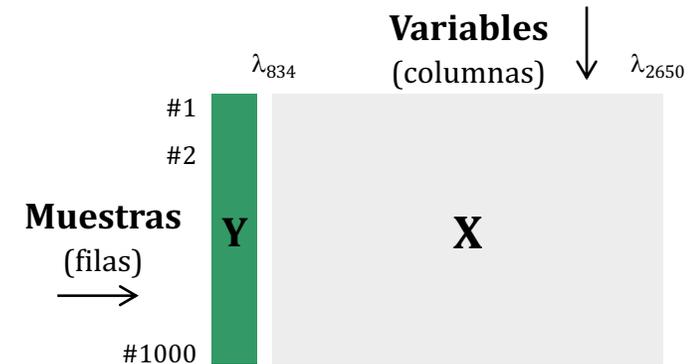
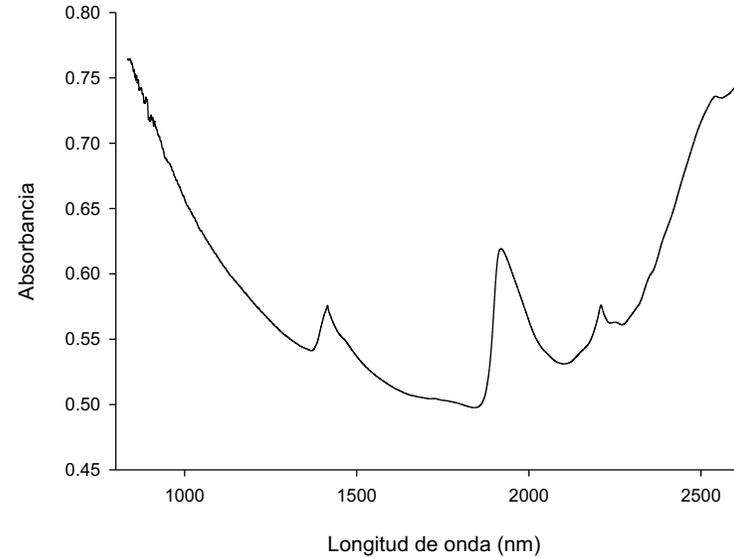
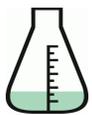
- **Combinación de bandas:**  
 $1691 + 3530 = 5221 \text{ cm}^{-1}$  (~1915 nm, NIR)
- **Primer sobretono (1st overtone):**  
 $2 * (3530) = 7060 \text{ cm}^{-1}$  (~1416 nm, NIR)
- **Segundo sobretono (2nd overtone):**  
 $3 * (3530) = 10590 \text{ cm}^{-1}$  (~944 nm, NIR)

Matriz NIR (información espectral)

		$\lambda_{834}$	$\lambda_{835}$	$\lambda_{836}$	$\lambda_{837}$	$\lambda_{838}$	....	$\lambda_{2649}$	$\lambda_{2650}$
Muestra1	$Y_1$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	....	$X_{1,p-1}$	$X_{1p}$
Muestra2	$Y_2$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	....	$X_{2,p-1}$	$X_{2p}$
Muestra 3	$Y_3$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	....	$X_{3,p-1}$	$X_{3p}$
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Muestra $n$	$Y_n$	$X_{n,1}$	$X_{n,2}$	$X_{n,3}$	$X_{n,4}$	$X_{n,5}$	....	$X_{n,p-1}$	$X_{np}$

Variable dependiente (Y)

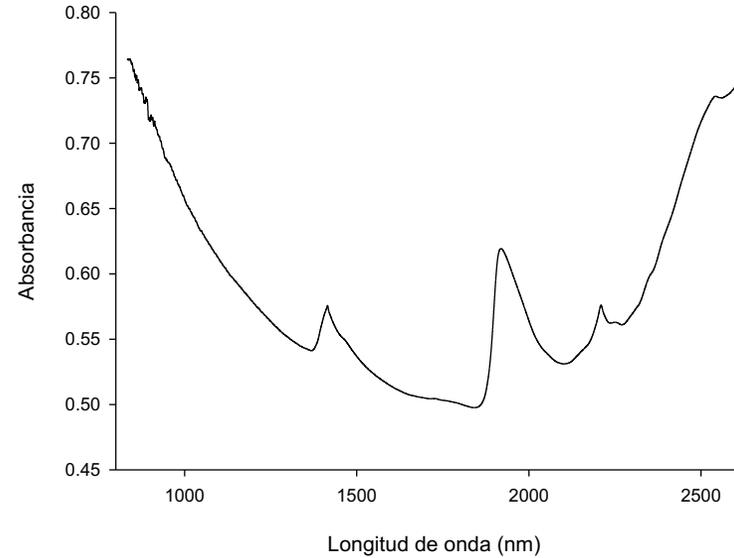
Variables independientes (X)



Ejemplo: si tenemos valores de absorbancia desde 834 nm hasta 2650 nm (NIR), a cada nanómetro (1 nm), tendremos 1816 valores ( $2650-834=1816$ ) por espectro, o 1816 variables 'independientes' variables. Si estamos trabajando con 1000 muestras (1000 espectros), tendremos una gran matriz de 1,816.000 celdas.

Matriz NIR (información espectral)

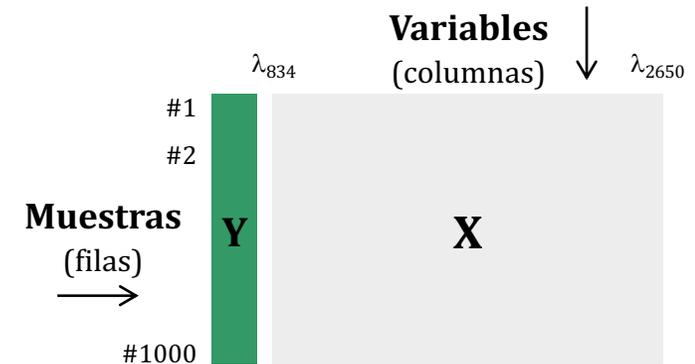
		$\lambda_{834}$	$\lambda_{835}$	$\lambda_{836}$	$\lambda_{837}$	$\lambda_{838}$	....	$\lambda_{2649}$	$\lambda_{2650}$
Muestra1	$CO_1$	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	....	$X_{1,p-1}$	$X_{1p}$
Muestra2	$CO_2$	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	....	$X_{2,p-1}$	$X_{2p}$
Muestra 3	$CO_3$	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	....	$X_{3,p-1}$	$X_{3p}$
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Muestra n	$CO_n$	$X_{n,1}$	$X_{n,2}$	$X_{n,3}$	$X_{n,4}$	$X_{n,5}$	....	$X_{n,p-1}$	$X_{np}$



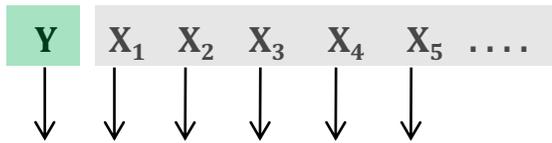
Carbono orgánico (Y)



Variables independientes (X)



Ejemplo: si tenemos valores de absorbancia desde 834 nm hasta 2650 nm (NIR), a cada nanómetro (1 nm), tendremos 1816 valores ( $2650-834=1816$ ) por espectro, o 1816 variables 'independientes' variables. Si estamos trabajando con 1000 muestras (1000 espectros), tendremos una gran matriz de 1,816.000 celdas.

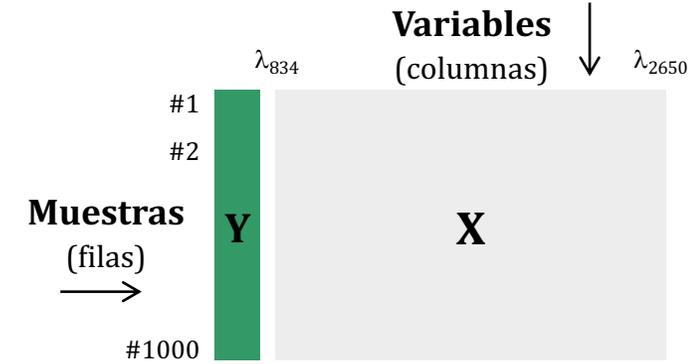
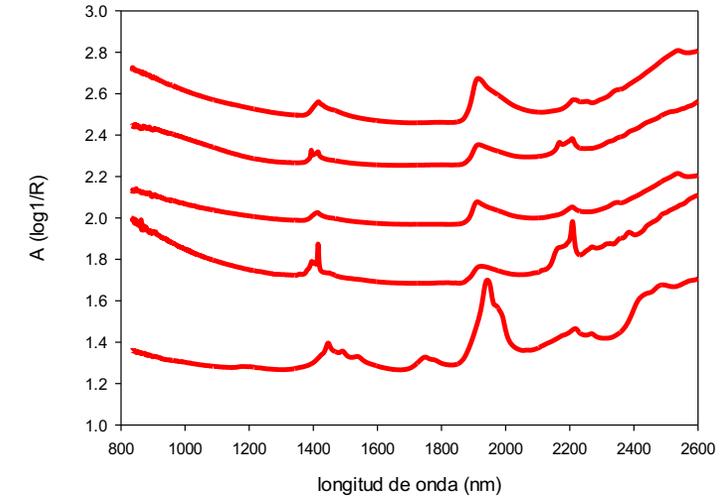


Variables

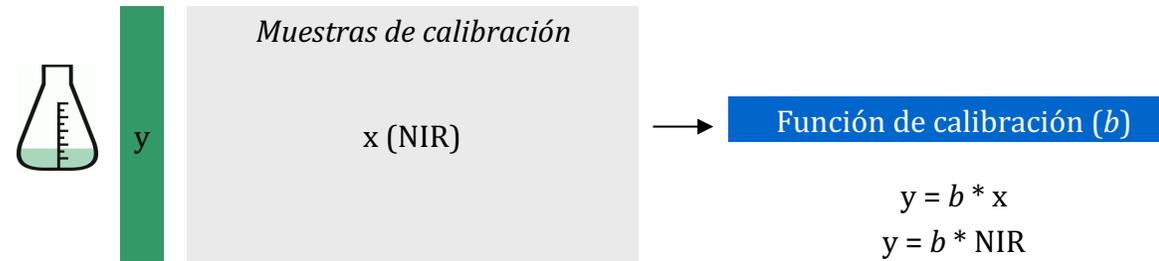
Muestras



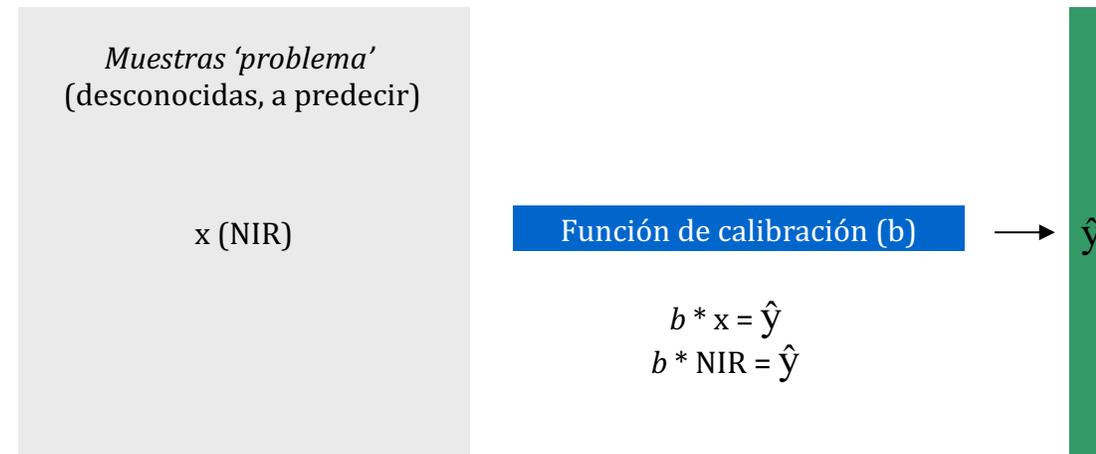
		834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	
	SOC	1'	2'	3'	4'	5'	6'	7'	8'	9'	10'	11'	12'	13'	14'	15'	16'	17'	18'	19'
01.0	1	0.7807	0.7639	0.7632	0.7638	0.7648	0.7673	0.7639	0.7614	0.7597	0.7638	0.7679	0.7613	0.7641	0.7643	0.7675	0.7660	0.7584	0.7595	0.7591
02.0	2	1.6248	0.8263	0.8228	0.8265	0.8151	0.8259	0.8270	0.8207	0.8243	0.8178	0.8186	0.8242	0.8254	0.8173	0.8167	0.8205	0.8280	0.8207	0.8151
03.0	3	2.0622	0.7876	0.7886	0.8003	0.7903	0.8041	0.7941	0.7959	0.7916	0.7961	0.7992	0.7992	0.7986	0.7964	0.7922	0.7979	0.7964	0.7911	0.7921
04.0	4	1.0371	0.6635	0.6644	0.6652	0.6631	0.6591	0.6642	0.6639	0.6631	0.6631	0.6629	0.6628	0.6574	0.6588	0.6606	0.6596	0.6555	0.6574	0.6541
05.0	5	1.1595	0.6253	0.6234	0.6240	0.6263	0.6278	0.6269	0.6229	0.6218	0.6237	0.6215	0.6198	0.6178	0.6214	0.6211	0.6235	0.6231	0.6200	0.6151
06.0	6	1.0102	0.6757	0.6730	0.6767	0.6782	0.6761	0.6731	0.6744	0.6723	0.6695	0.6728	0.6689	0.6692	0.6715	0.6654	0.6736	0.6661	0.6714	0.6701
07.0	7	1.7030	0.8072	0.8096	0.8088	0.8066	0.8016	0.8145	0.8068	0.8017	0.7994	0.8118	0.8066	0.8035	0.8005	0.8015	0.7952	0.8018	0.7965	0.7921
08.0	8	1.0169	0.7531	0.7506	0.7589	0.7534	0.7558	0.7474	0.7562	0.7542	0.7551	0.7600	0.7541	0.7559	0.7588	0.7508	0.7479	0.7448	0.7442	0.7501
09.0	9	0.5730	0.7221	0.7188	0.7233	0.7338	0.7295	0.7296	0.7239	0.7252	0.7245	0.7206	0.7225	0.7229	0.7218	0.7235	0.7194	0.7244	0.7182	0.7181
10.0	10	0.9447	0.7086	0.7106	0.7084	0.7075	0.7062	0.7067	0.7086	0.7056	0.7042	0.7060	0.6939	0.7024	0.7035	0.7008	0.7024	0.7088	0.6974	0.6951
11.0	11	0.7798	0.7231	0.7241	0.7290	0.7298	0.7261	0.7291	0.7265	0.7240	0.7178	0.7253	0.7177	0.7190	0.7208	0.7274	0.7310	0.7245	0.7239	0.7201
12.0	12	2.6900	0.8169	0.8161	0.8187	0.8168	0.8114	0.8181	0.8095	0.8028	0.8060	0.8092	0.8162	0.8123	0.8172	0.8187	0.8187	0.8184	0.8137	0.8061
13.0	13	1.5360	0.7935	0.7901	0.7879	0.7815	0.7876	0.7848	0.7893	0.7845	0.7916	0.7849	0.7815	0.7885	0.7866	0.7837	0.7858	0.7848	0.7871	0.7851
14.0	14	1.6891	0.8040	0.8003	0.8022	0.7987	0.8033	0.8022	0.8067	0.8013	0.7998	0.8068	0.7949	0.8007	0.8015	0.8035	0.8000	0.7967	0.7991	0.7931
15.0	15	1.1904	0.7808	0.7826	0.7809	0.7789	0.7845	0.7836	0.7808	0.7774	0.7723	0.7809	0.7775	0.7809	0.7761	0.7813	0.7859	0.7755	0.7782	0.7761
16.0	16	3.1769	0.8185	0.8222	0.8183	0.8152	0.8208	0.8137	0.8168	0.8195	0.8195	0.8110	0.8176	0.8203	0.8192	0.8091	0.8126	0.8210	0.8149	0.8141
17.0	17	2.4594	0.8289	0.8320	0.8321	0.8318	0.8353	0.8347	0.8288	0.8306	0.8320	0.8305	0.8372	0.8315	0.8333	0.8334	0.8348	0.8321	0.8365	0.8331
18.0	18	1.5953	0.8031	0.8073	0.7997	0.7998	0.8073	0.8070	0.8023	0.8076	0.8049	0.8048	0.8088	0.8084	0.8008	0.8028	0.8052	0.7994	0.8030	0.8001
19.0	19	3.1097	0.8199	0.8181	0.8269	0.8247	0.8216	0.8170	0.8173	0.8262	0.8232	0.8217	0.8168	0.8176	0.8181	0.8190	0.8225	0.8188	0.8137	0.8151
20.0	20	0.8504	0.8443	0.8311	0.8318	0.8419	0.8414	0.8370	0.8385	0.8387	0.8381	0.8416	0.8413	0.8431	0.8335	0.8319	0.8343	0.8354	0.8291	0.8261
21.0	21	1.0012	0.7891	0.7933	0.7926	0.7874	0.7926	0.7867	0.7939	0.7854	0.7974	0.7934	0.7906	0.7916	0.7863	0.7832	0.7948	0.7868	0.7872	0.7851
22.0	22	1.0614	0.8169	0.8165	0.8160	0.8282	0.8268	0.8263	0.8208	0.8228	0.8161	0.8220	0.8184	0.8136	0.8143	0.8172	0.8175	0.8138	0.8089	0.8121
23.0	23	1.8997	0.8126	0.8141	0.8093	0.8102	0.8074	0.8094	0.8040	0.8043	0.8024	0.8092	0.8075	0.8046	0.8052	0.7994	0.7988	0.8045	0.8041	0.7971
24.0	24	1.7809	0.8445	0.8462	0.8372	0.8385	0.8444	0.8413	0.8477	0.8413	0.8446	0.8508	0.8402	0.8346	0.8432	0.8367	0.8482	0.8445	0.8449	0.8351
25.0	25	1.0967	0.8094	0.8176	0.8226	0.8175	0.8167	0.8128	0.8149	0.8273	0.8199	0.8159	0.8100	0.8131	0.8124	0.8166	0.8181	0.8194	0.8145	0.8101
26.0	26	3.7051	0.8650	0.8623	0.8601	0.8628	0.8580	0.8624	0.8634	0.8638	0.8546	0.8707	0.8477	0.8488	0.8596	0.8485	0.8525	0.8594	0.8526	0.8521
27.0	27	1.7872	0.8162	0.8167	0.8138	0.8158	0.8089	0.8155	0.8109	0.8112	0.8153	0.8126	0.8139	0.8126	0.8131	0.8115	0.8067	0.8195	0.8112	0.8061
28.0	28	1.2748	0.8212	0.8184	0.8189	0.8210	0.8116	0.8146	0.8159	0.8199	0.8188	0.8206	0.8186	0.8159	0.8155	0.8134	0.8209	0.8181	0.8103	0.8151
29.0	29	1.0121	0.8161	0.8160	0.8170	0.8130	0.8147	0.8166	0.8180	0.8190	0.8174	0.8201	0.8165	0.8200	0.8200	0.8200	0.8200	0.8200	0.8200	0.8200



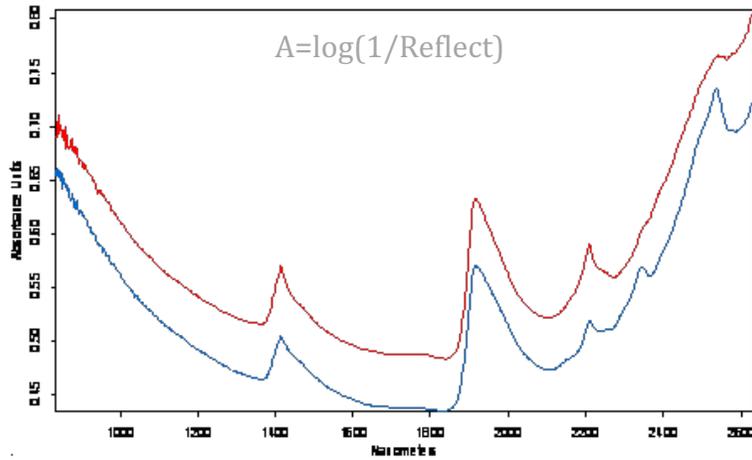
## 1º) Desarrollo de la calibración (o modelo)



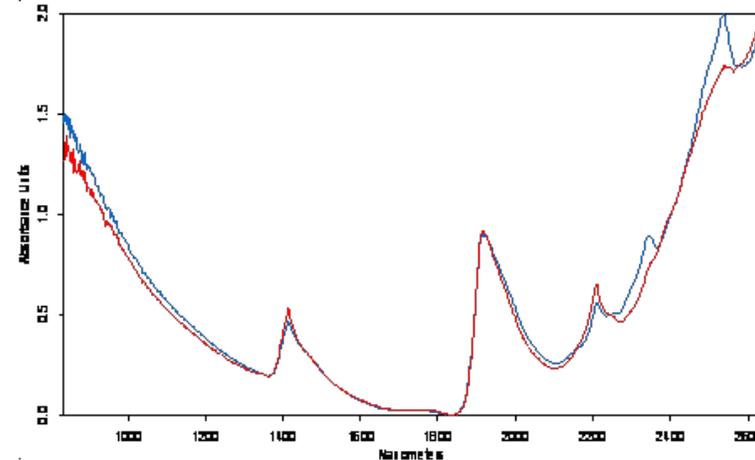
## 2º) Uso de la calibración



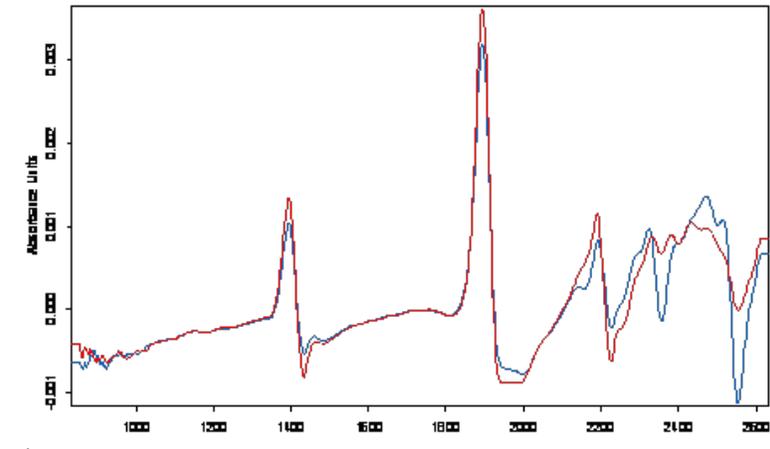
Sin preprocesamiento



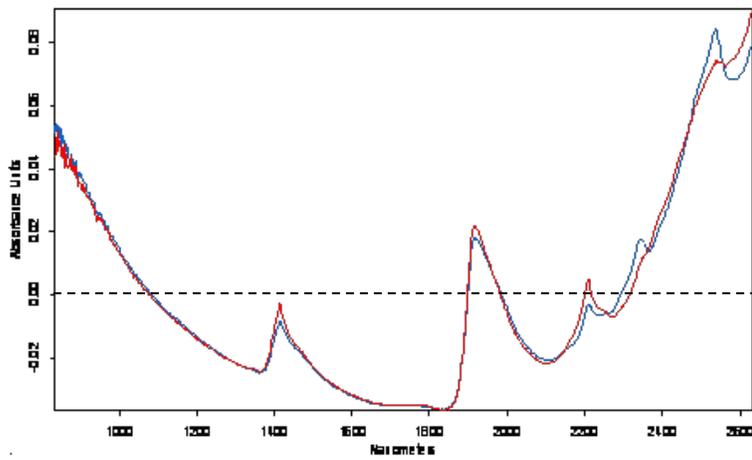
Normalización de máximos y mínimos



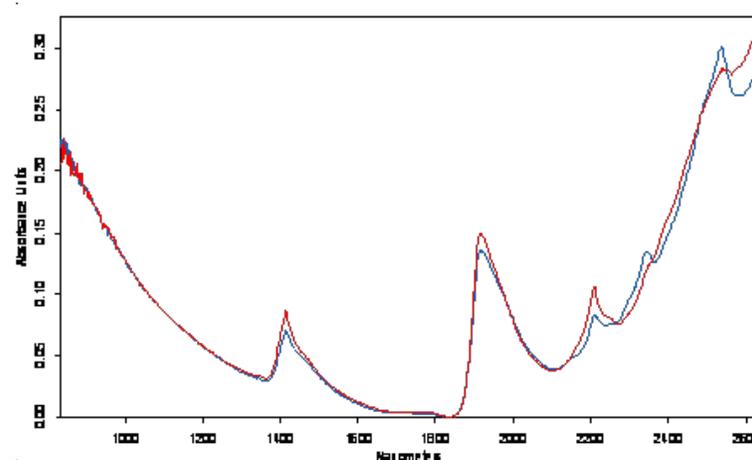
Primera derivada (Savitzky Golay, 25 pts.)



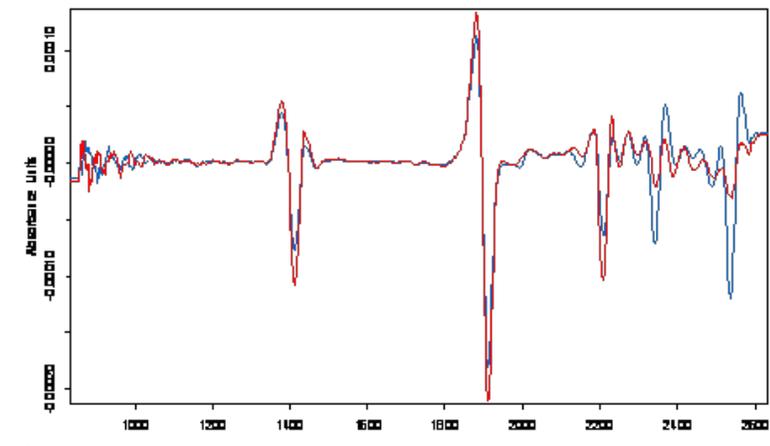
Standard Normal Variate (SNV)



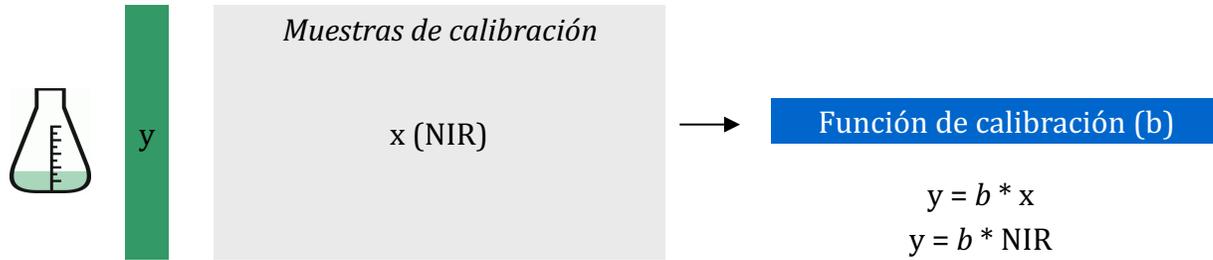
Línea base: Baseline offset



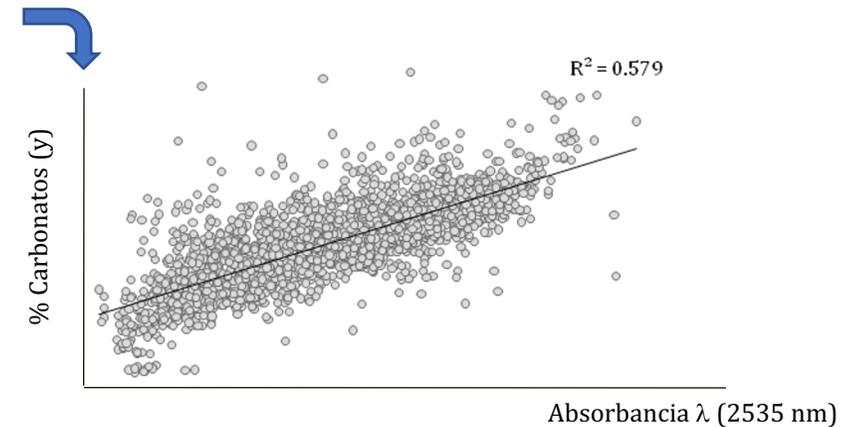
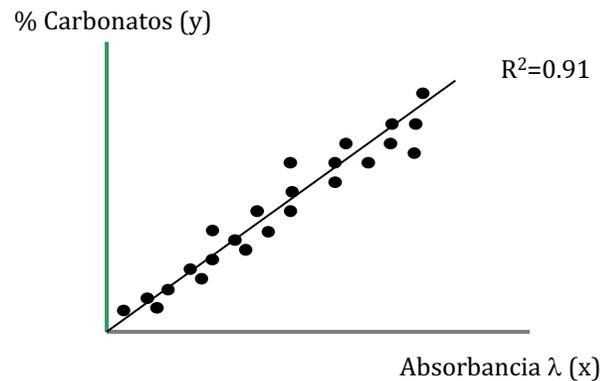
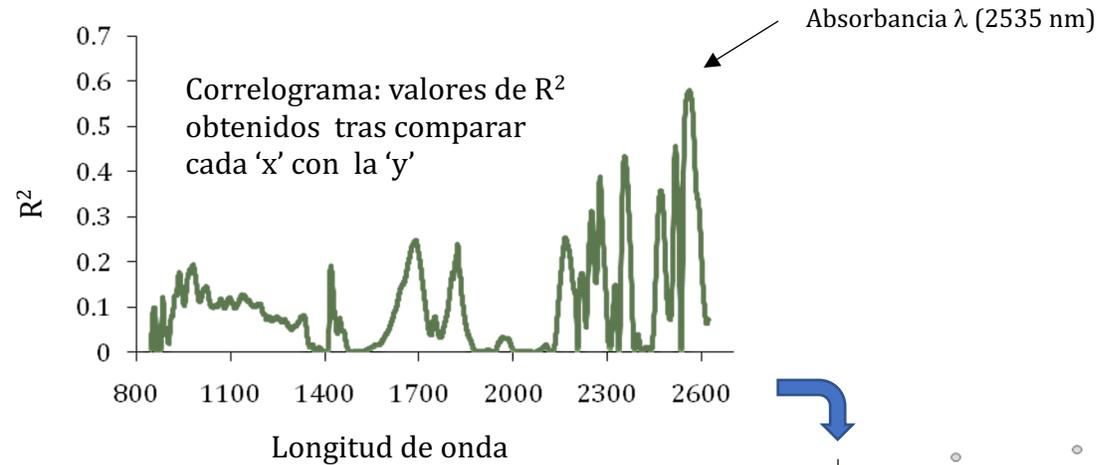
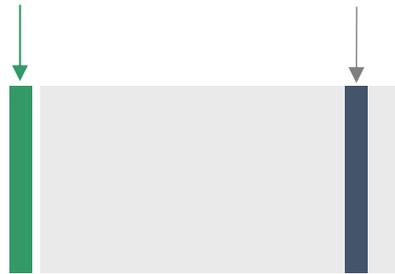
Segunda derivada (Savitzky Golay, 25 pts.)



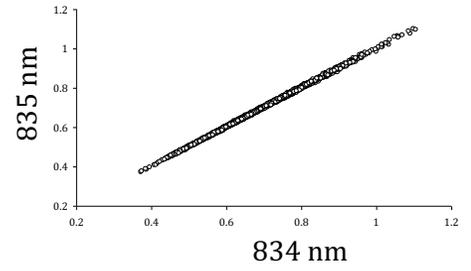
## 1) Desarrollo de la calibración



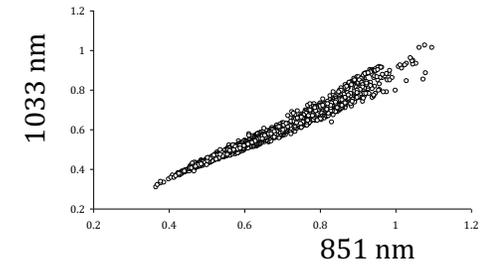
\* Regresión lineal univariante



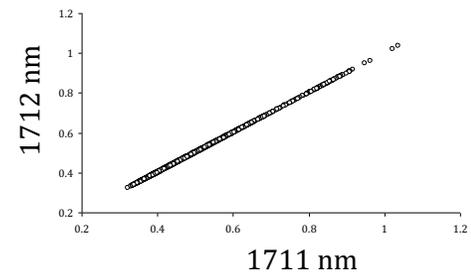
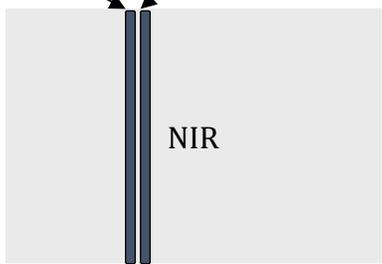
834 nm 835 nm



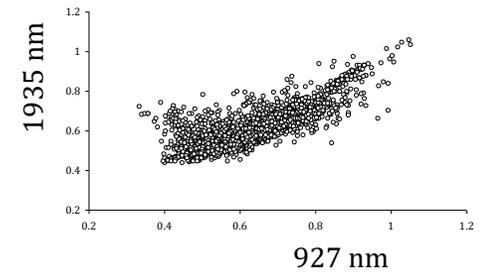
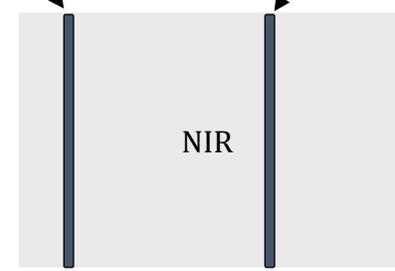
851 nm 1033 nm



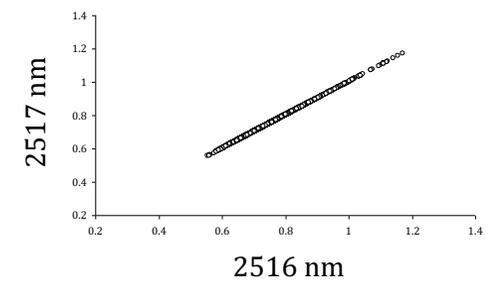
1711 nm 1712 nm



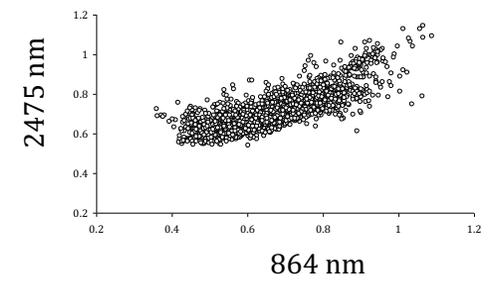
927 nm 1935 nm



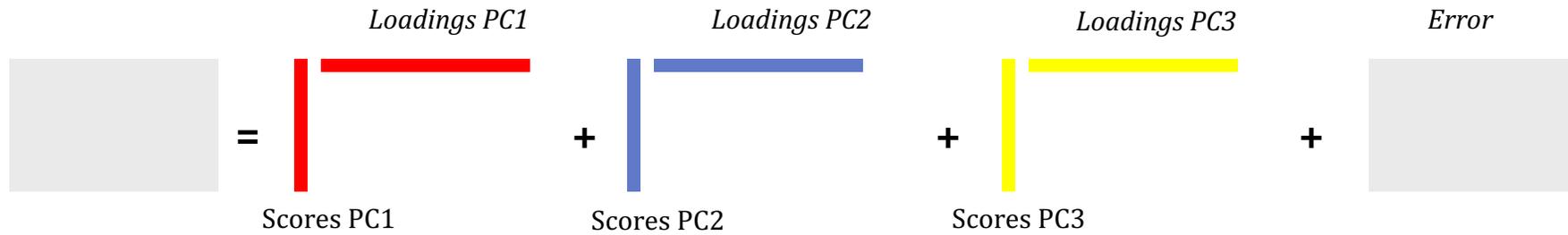
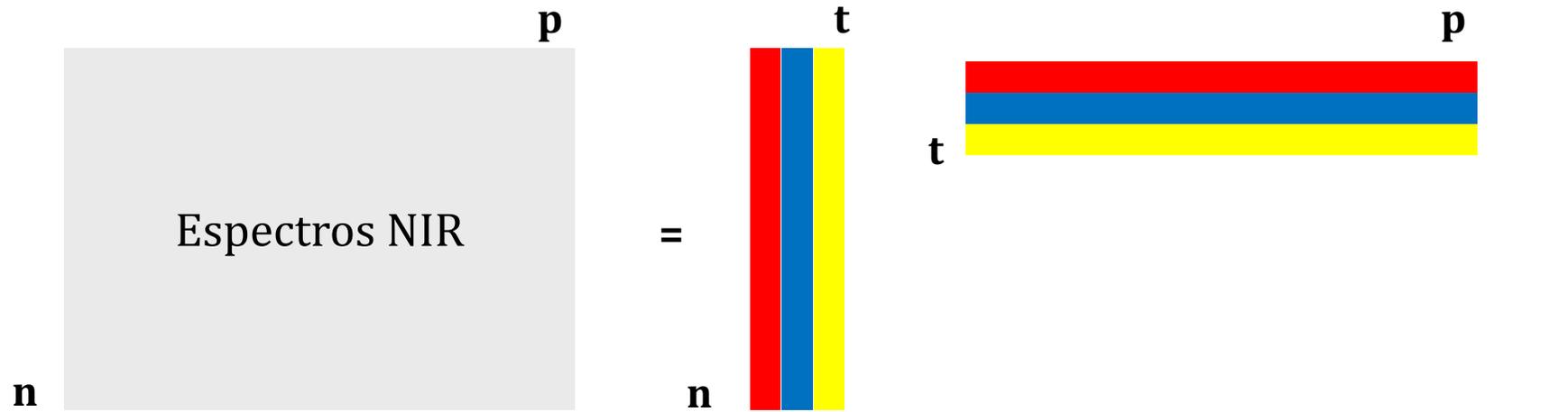
2516 nm 2517 nm



864 nm 2475 nm



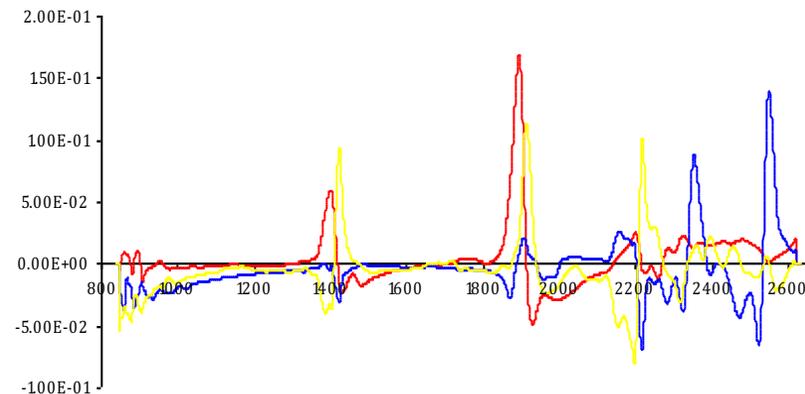
# Análisis de componentes principales (PCA)



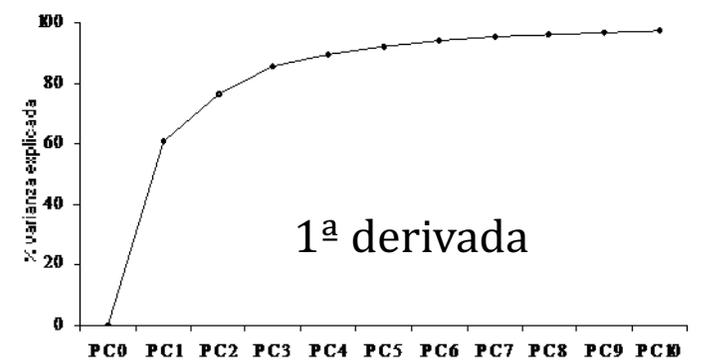
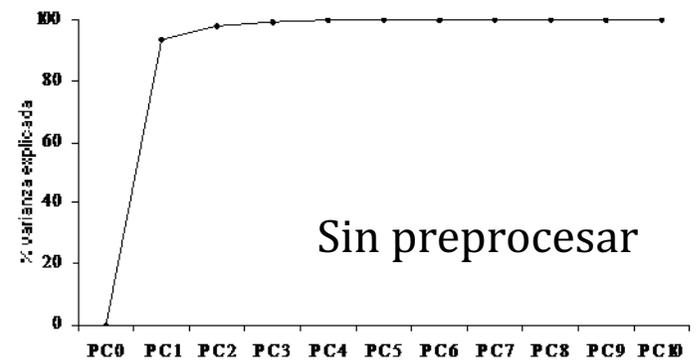
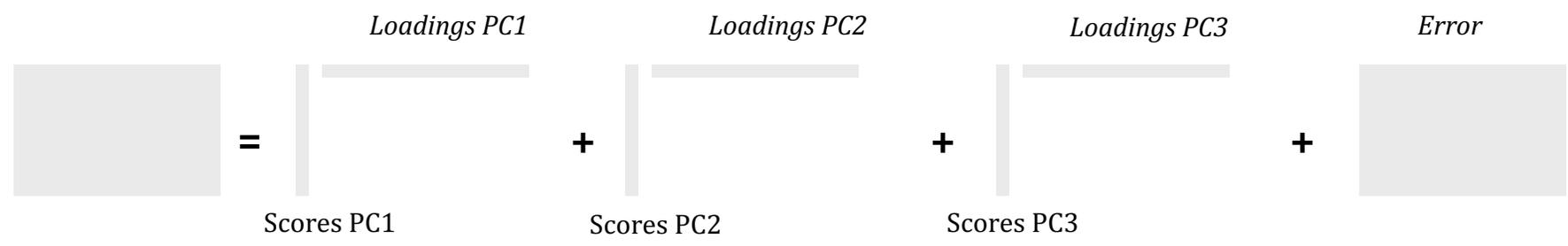
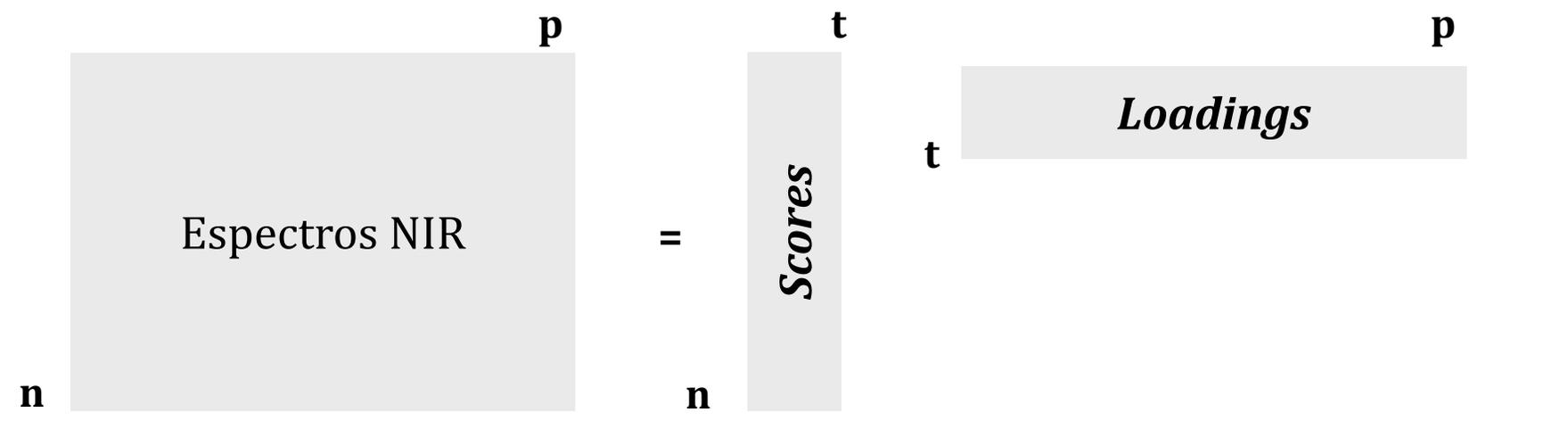
$$\text{PC1} = L_{1,\text{PC1}} \times \lambda_1 + L_{2,\text{PC1}} \times \lambda_2 + L_{3,\text{PC1}} \times \lambda_3 \dots + L_{p,\text{PC1}} \times \lambda_p$$

$$\text{PC2} = L_{1,\text{PC2}} \times \lambda_1 + L_{2,\text{PC2}} \times \lambda_2 + L_{3,\text{PC2}} \times \lambda_3 \dots + L_{p,\text{PC2}} \times \lambda_p$$

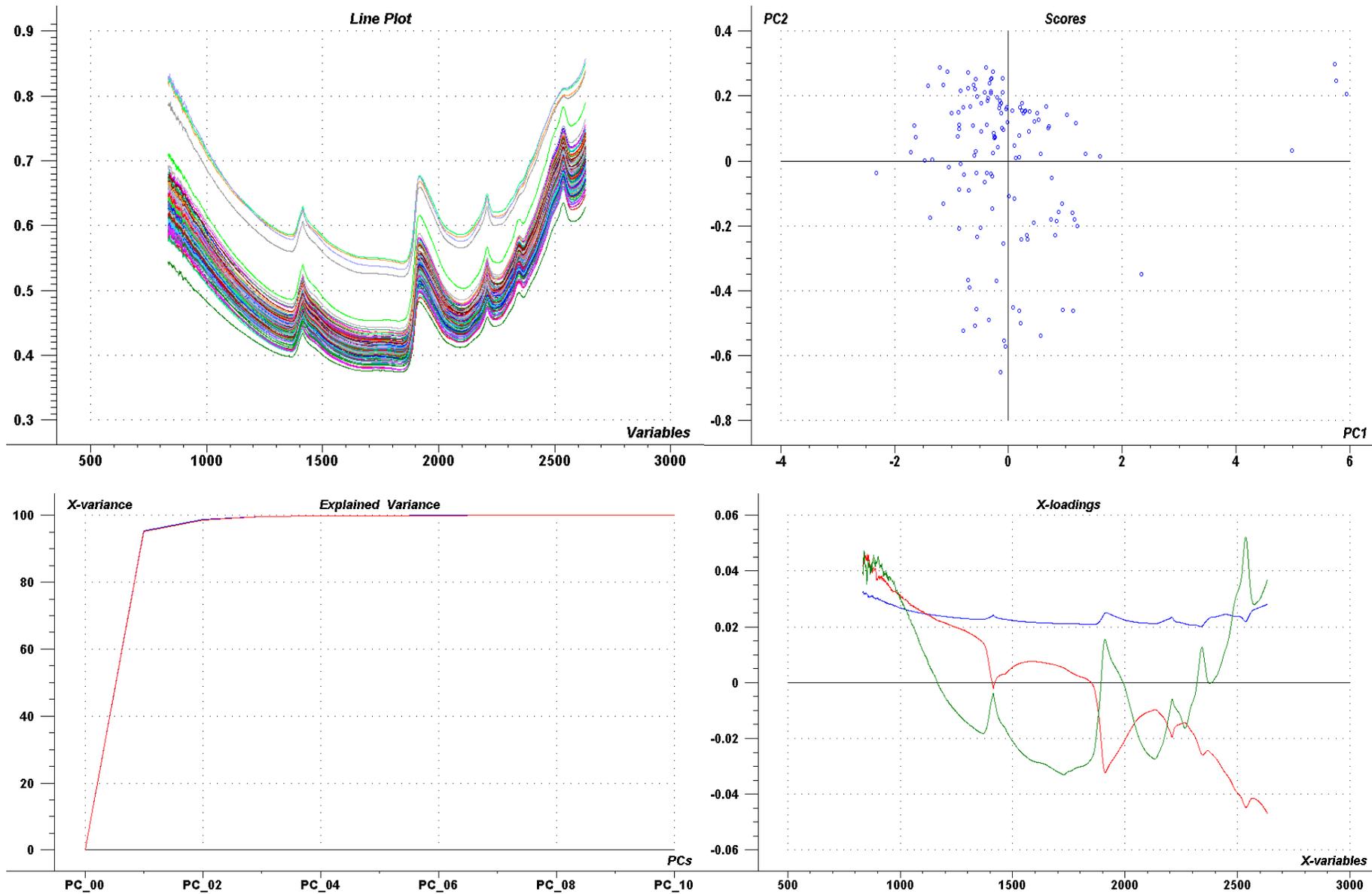
$$\text{PC3} = L_{1,\text{PC3}} \times \lambda_1 + L_{2,\text{PC3}} \times \lambda_2 + L_{3,\text{PC3}} \times \lambda_3 \dots + L_{p,\text{PC3}} \times \lambda_p$$



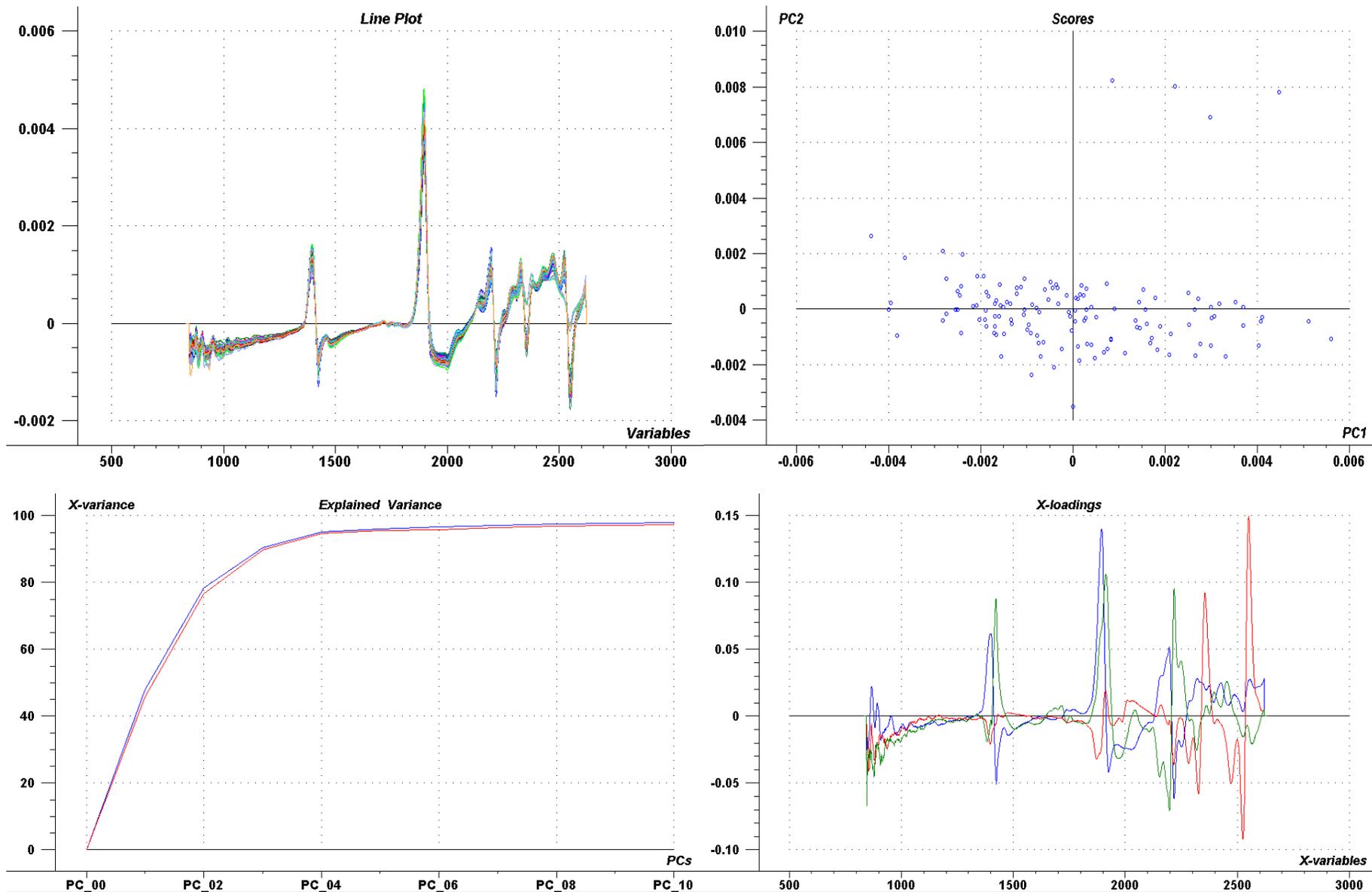
# Análisis de componentes principales (PCA)



# Análisis de componentes principales (PCA)



# Análisis de componentes principales (PCA)



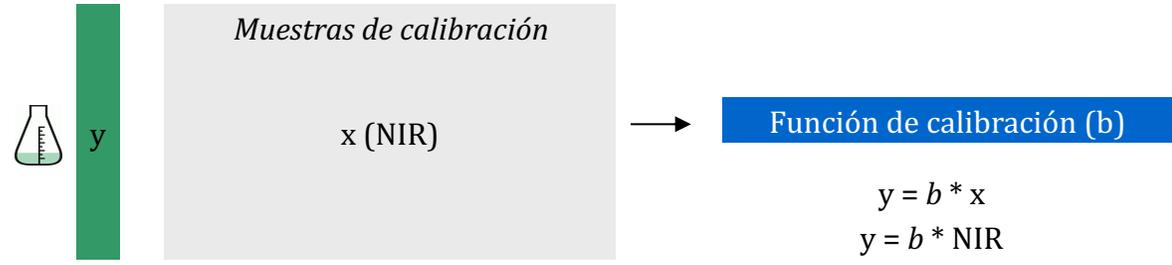
	Matriz de espectros NIR					Matriz de puntuaciones		
	$\lambda_1$	$\lambda_2$	$\lambda_3$	...	..... $\lambda_p$	<b>PC<sub>1</sub></b>	<b>PC<sub>2</sub></b>	<b>PC<sub>3</sub></b>
Sample 1	$X_{s1,\lambda_1}$	$X_{s1,\lambda_2}$	$X_{s1,\lambda_3}$	...	..... $X_{s1,\lambda_p}$	$X_{S1,PC1}$	$X_{S1,PC2}$	$X_{S1,PC3}$
Sample 2	$X_{s2,\lambda_1}$	$X_{s2,\lambda_2}$	$X_{s2,\lambda_3}$	...	..... $X_{s2,\lambda_p}$	$X_{S2,PC1}$	$X_{S2,PC2}$	$X_{S2,PC3}$
Sample 3	$X_{s3,\lambda_1}$	$X_{s3,\lambda_2}$	$X_{s3,\lambda_3}$	...	..... $X_{s3,\lambda_p}$	$X_{S3,PC1}$	$X_{S3,PC2}$	$X_{S3,PC3}$
Sample 4	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
Sample $n$	$X_{sn,\lambda_1}$	$X_{sn,\lambda_2}$	$X_{sn,\lambda_3}$	...	..... $X_{sn,\lambda_p}$	$X_{Sn,PC1}$	$X_{Sn,PC2}$	$X_{Sn,PC3}$
	Variables independientes (x)					Puntuaciones factoriales o <i>scores</i>		

$$PC1 = L_{1,PC1} \times \lambda_1 + L_{2,PC1} \times \lambda_2 + L_{3,PC1} \times \lambda_3 \dots + L_{p,PC1} \times \lambda_p$$

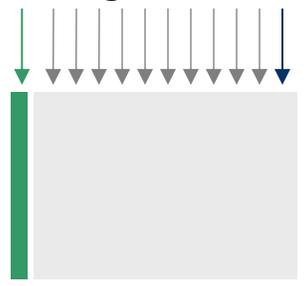
$$PC2 = L_{1,PC2} \times \lambda_1 + L_{2,PC2} \times \lambda_2 + L_{3,PC2} \times \lambda_3 \dots + L_{p,PC2} \times \lambda_p$$

$$PC3 = L_{1,PC3} \times \lambda_1 + L_{2,PC3} \times \lambda_2 + L_{3,PC3} \times \lambda_3 \dots + L_{p,PC3} \times \lambda_p$$

## 1) Desarrollo de la calibración



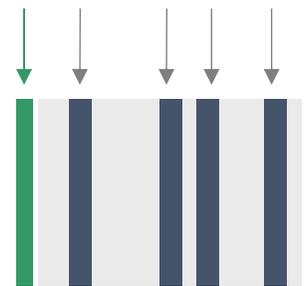
### \* Regresión lineal multivariante



MLR (*multiple linear regression*)

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 \dots + b_p \cdot x_p$$

Descartamos por colinealidad

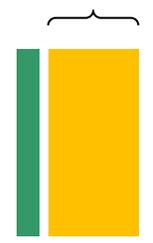


MLR *Stepwise*

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3$$

Útil (primeras calibraciones)

### PC-loadings

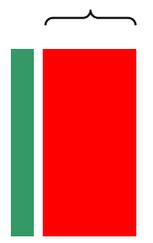


PCR  
(*Principal component regression*)

$$y = b_0 + b_1 \cdot \text{PC}_1 + b_2 \cdot \text{PC}_2 + b_3 \cdot \text{PC}_3 \dots$$

Los predictores son los *scores* obtenidos con los PC-loadings

### PLS-vectors

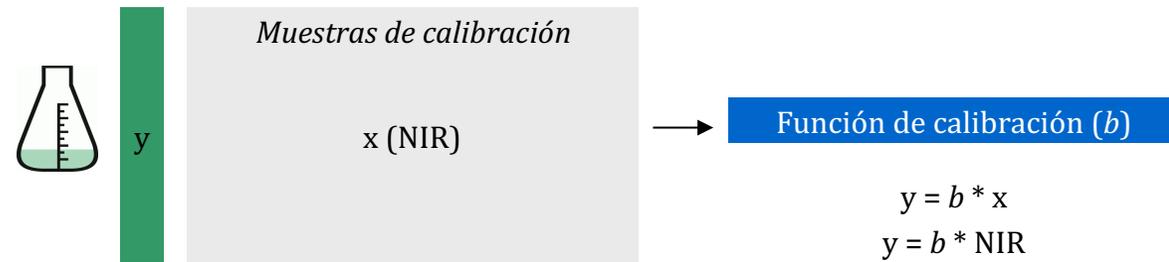


PLS  
(*partial least squares regression*)

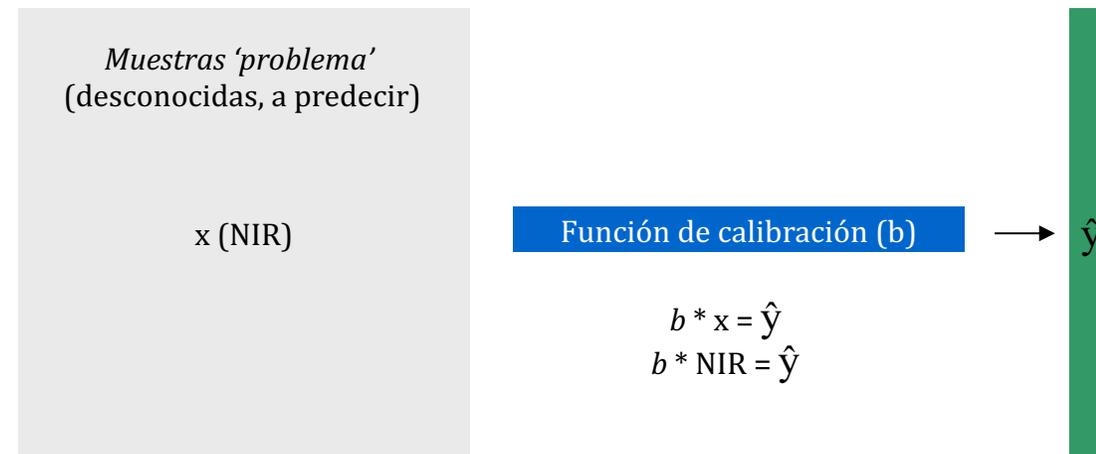
$$y = b_0 + b_1 \cdot \text{PLSv}_1 + b_2 \cdot \text{PLSv}_2 + b_3 \cdot \text{PLSv}_3 \dots$$

Los predictores son los *scores* obtenidos con los PLS-vectors

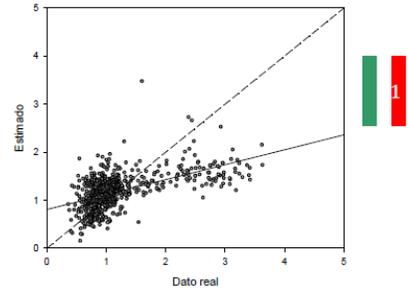
## 1º) Desarrollo de la calibración (o modelo)



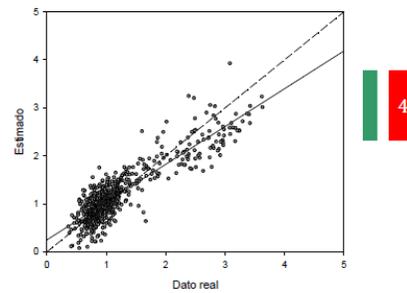
## 2º) Uso de la calibración



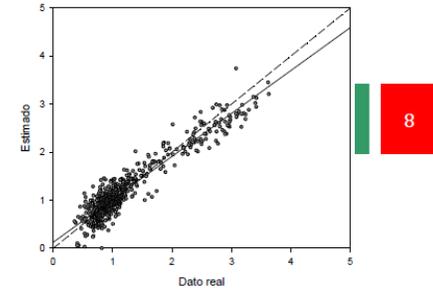
Rango 1



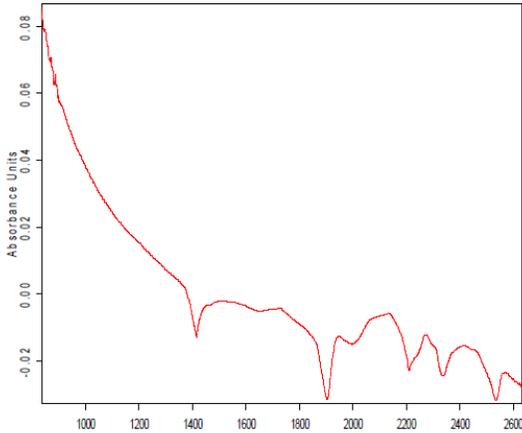
Rango 4



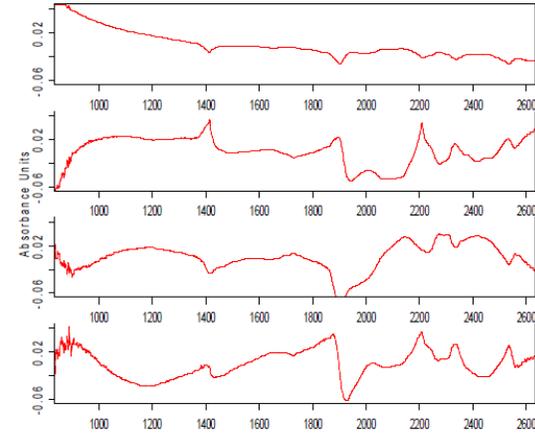
Rango 8



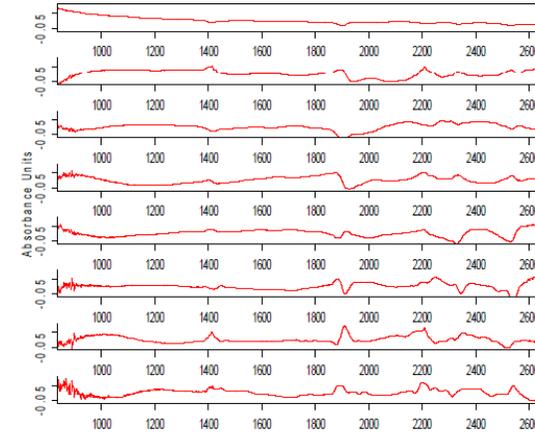
Loadings (PLS-vectors)



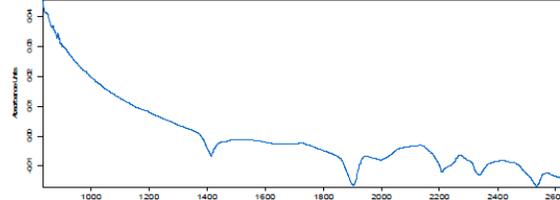
Loadings (PLS-vectors)



Loadings (PLS-vectors)



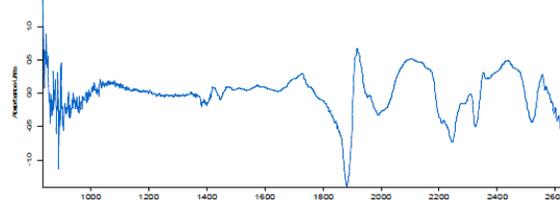
*b*-coefficients



*b*-coefficients



*b*-coefficients



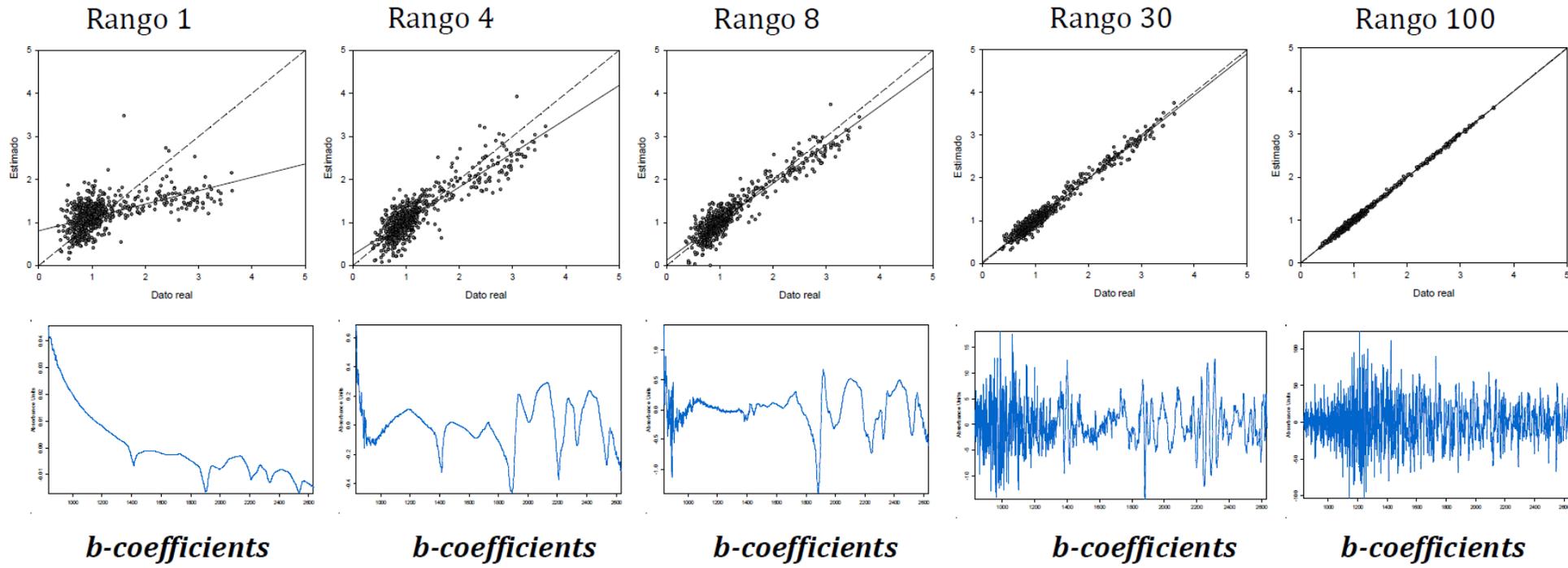
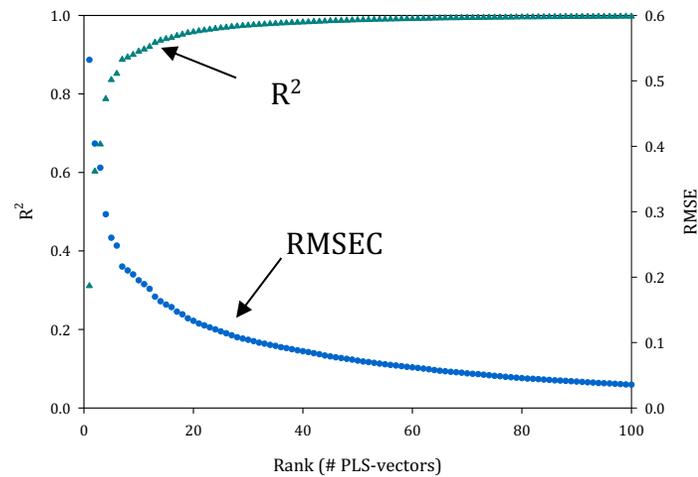
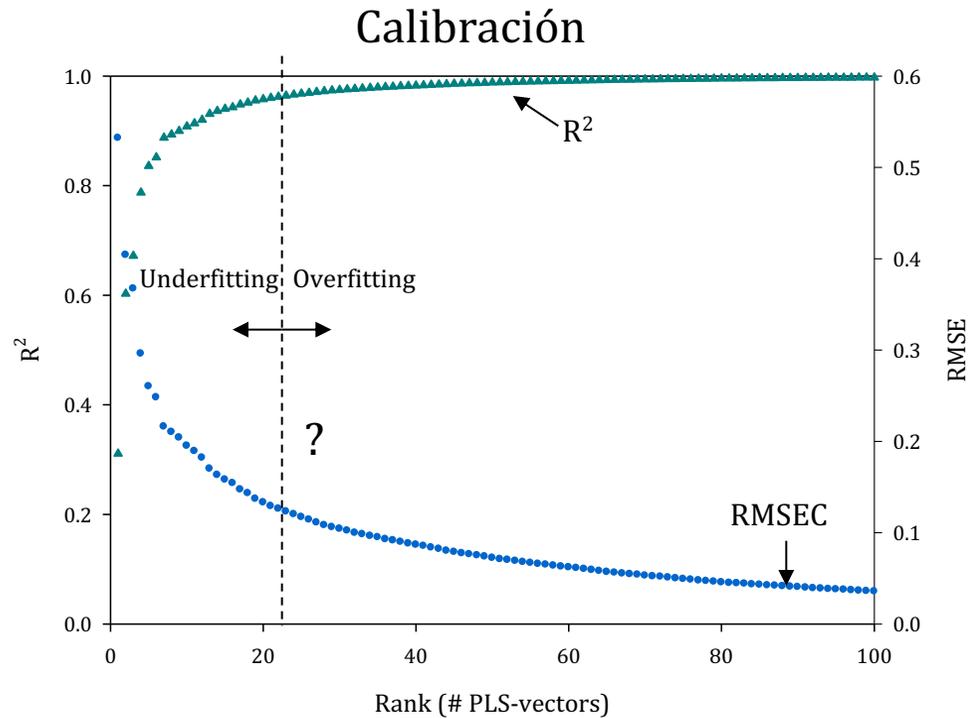


Gráfico de sedimentación  
(*scree plot*)

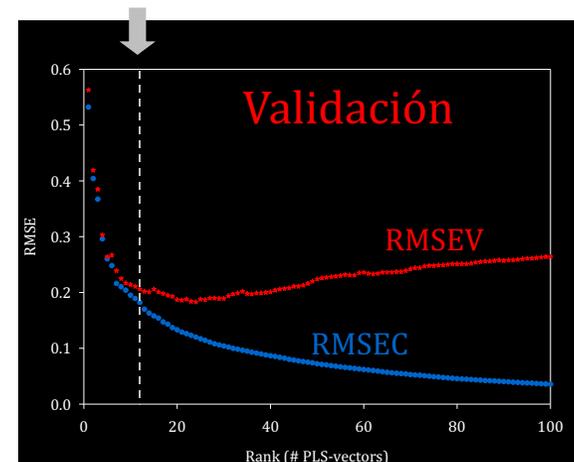
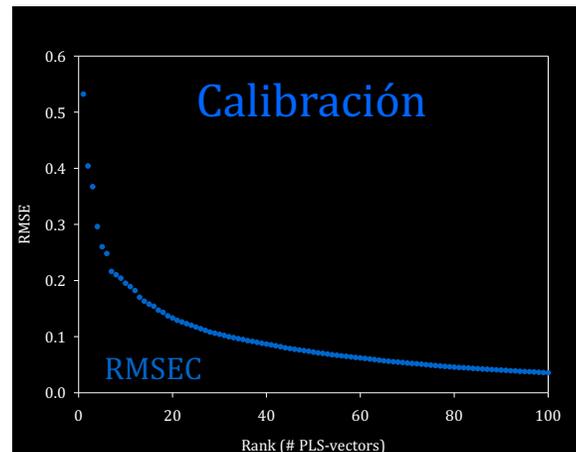


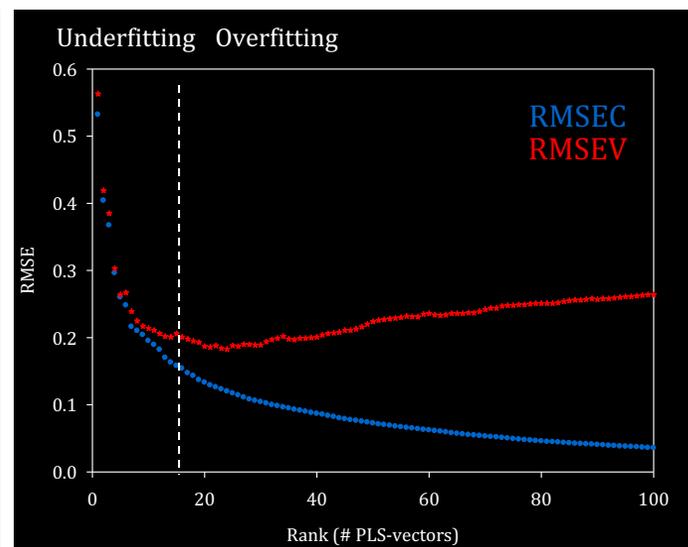
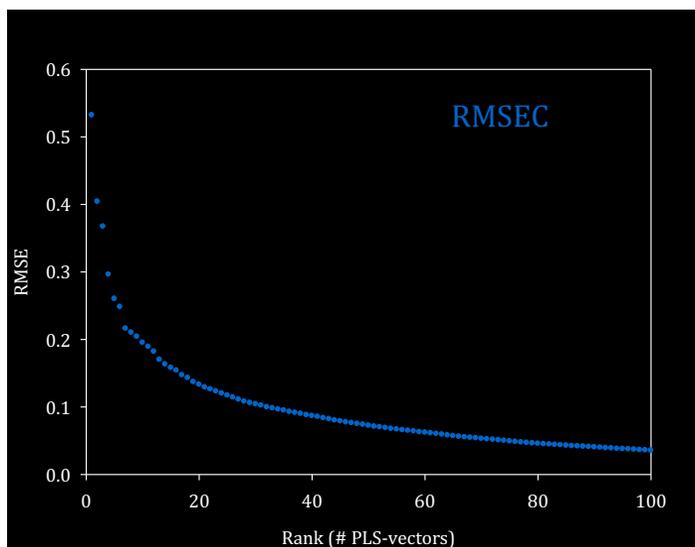
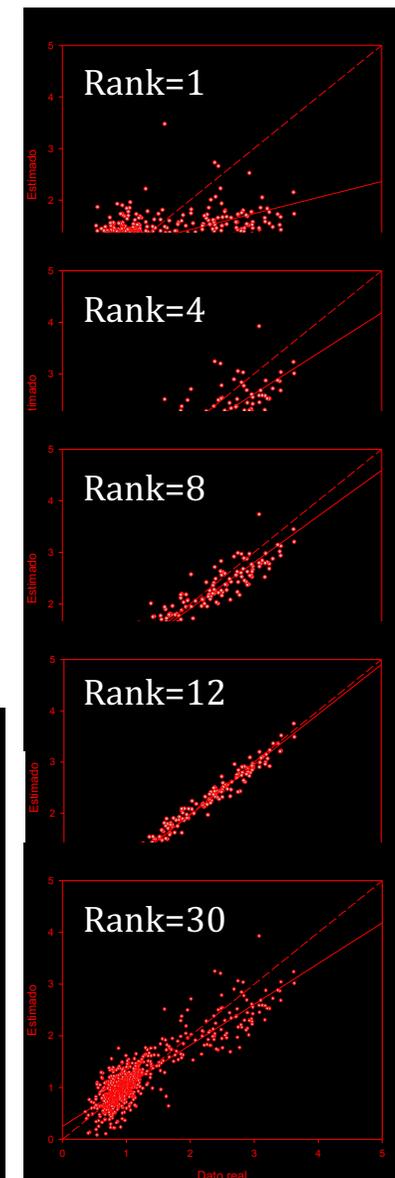
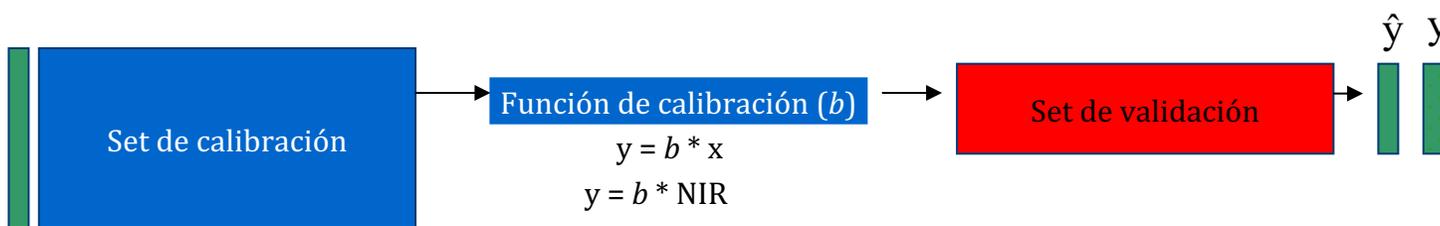
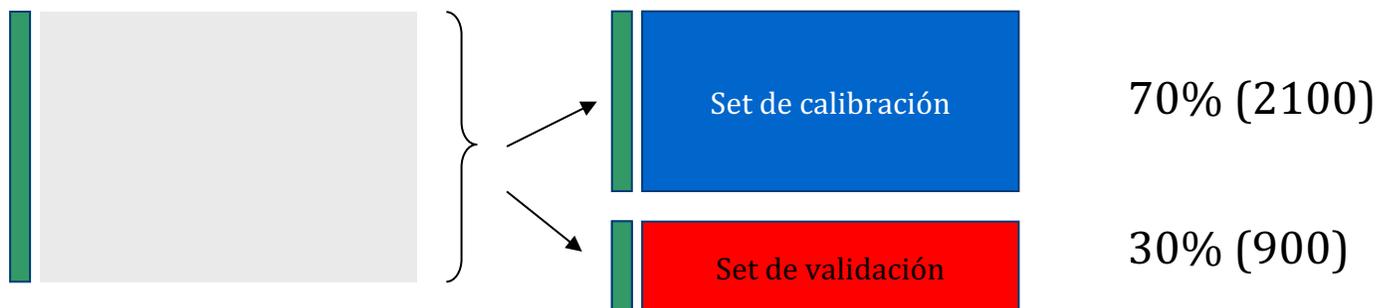
$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (Differ_i)^2}$$

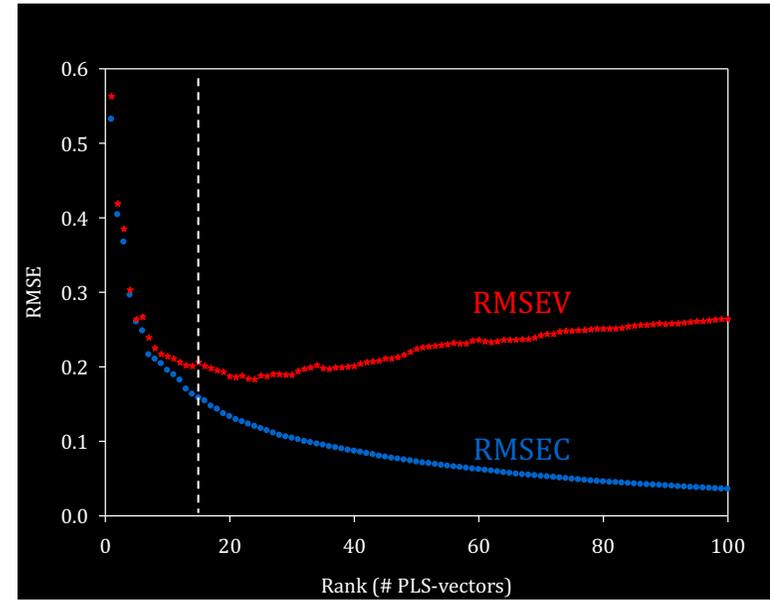
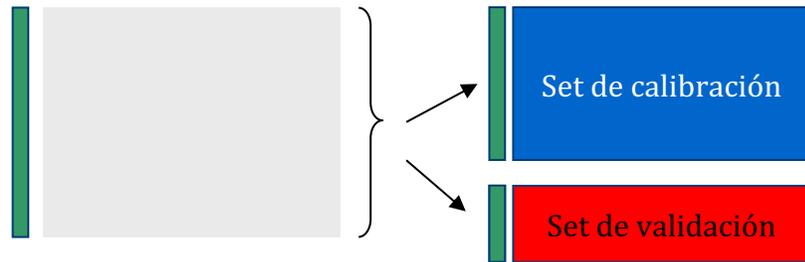
$Differ_i = \hat{y}_i \text{ estimado} - y_i \text{ real}$



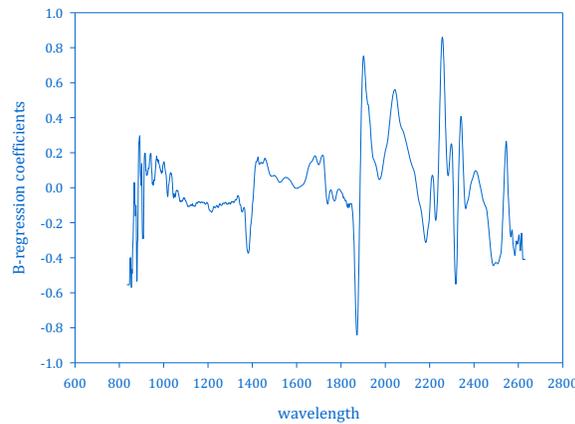
- La validación nos permite establecer **empíricamente** cuál es el rango óptimo (*rank*, **número de componentes**, o vectores-PLS), y por tanto delimitar entre dónde se produce “*underfitting*” y “*overfitting*”.
- Los primeros vectores-PLS son los que más información aportan (rango bajo), y a medida que aumentemos el rango, estaremos incluyendo cada vez más ruido dentro del modelo.
- Basándonos en los datos de la validación, también podemos decidir otros aspectos relacionados con la arquitectura del modelo, tales como el pre-procesamiento (derivadas, etc.) más adecuado.
- En realidad, más que una verdadera validación, lo que se está haciendo es **finalizar el ajuste de los parámetros del modelo** (*tuning*).
- Dos métodos: (a) *test-set validation* y (b) *cross-validation*





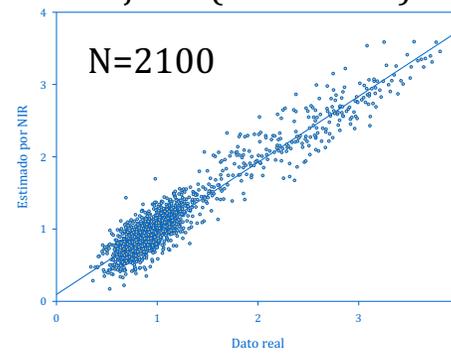


***b*-coefficients**

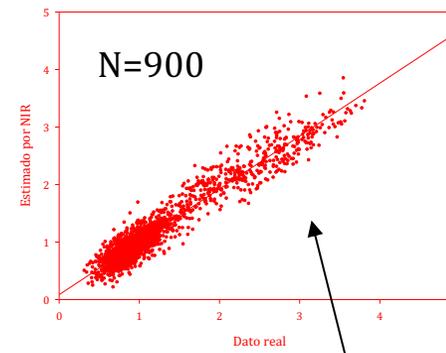


**Función de calibración**

**Ajuste (calibración)**

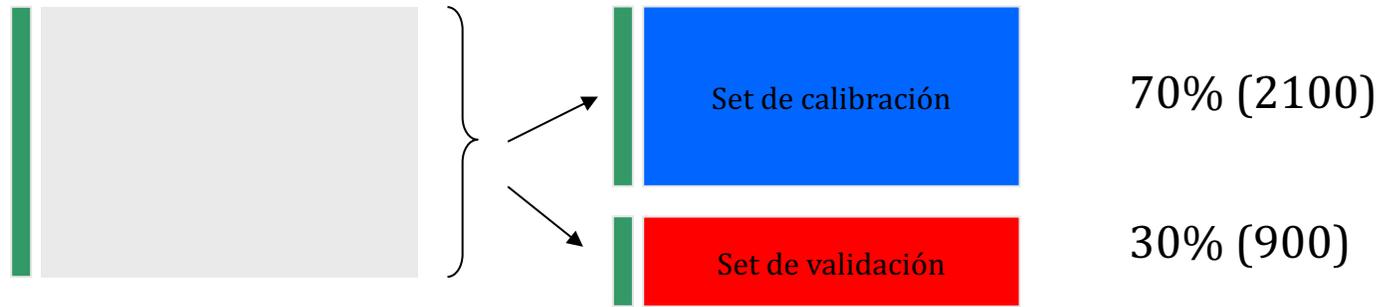


**Validación**



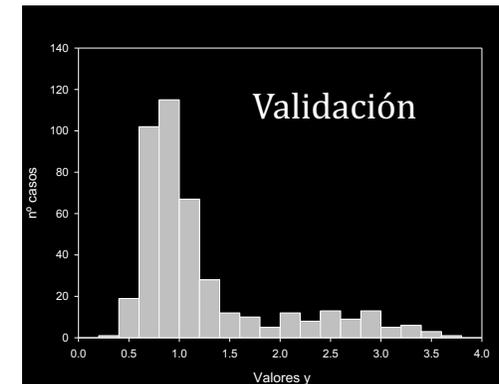
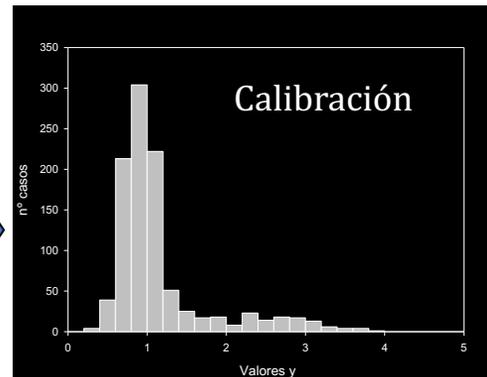
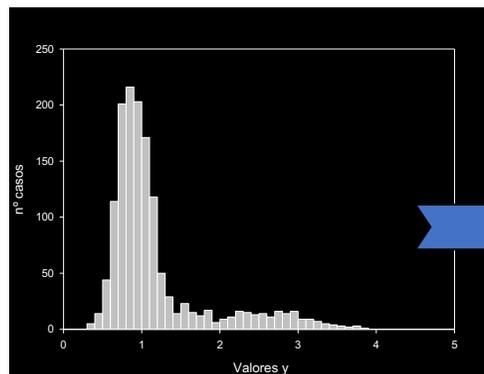
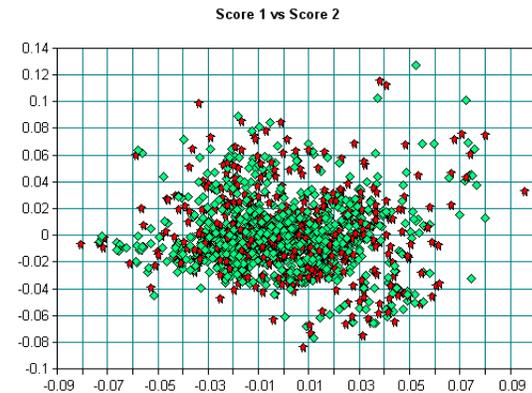
**$R^2$ , RMSEV, *bias*, SEP**

## La selección de muestras en *test set*



La selección de muestras en *test set*

- Kennard and Stone (1969) algorithm.*
- PCA y Clusters (Martens and Næs, 1989).*
- Rangos de la variable "y" similares*



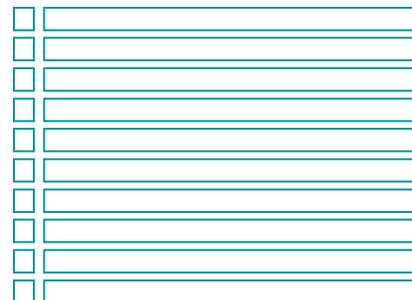


En “*cross-validation*” (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

Muestras calibración



Muestras validación





En "*cross-validation*" (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

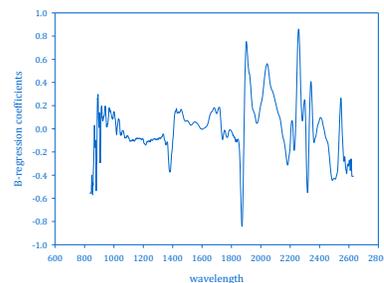
Muestras calibración

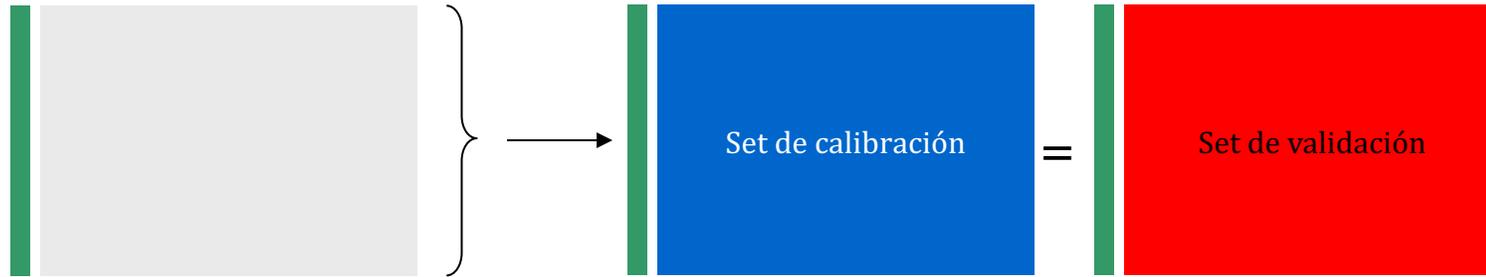


Muestras validación    Estimación  $\hat{y}$



Función de calibración ( $b$ )





En "*cross-validation*" (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

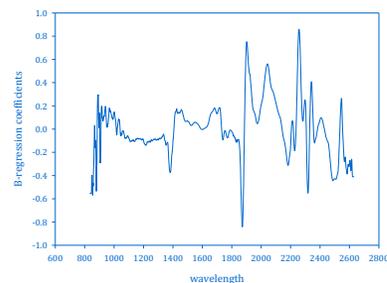
Muestras calibración

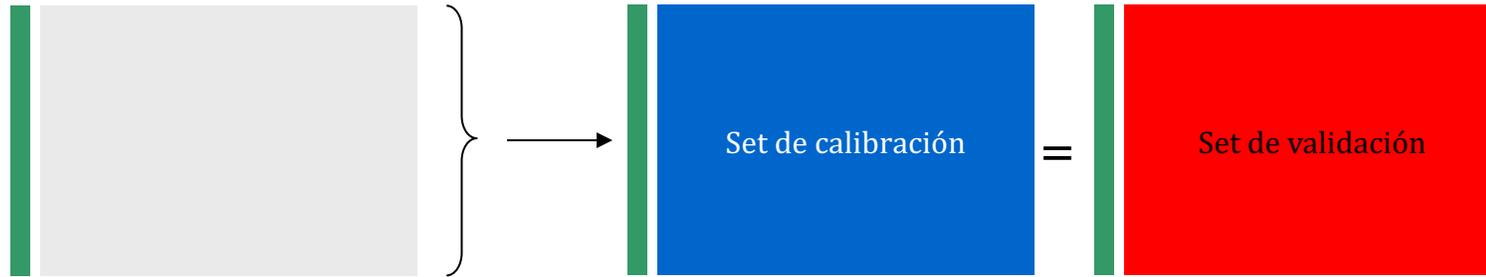


Muestras validación      Estimación  $\hat{y}$



Función de calibración ( $b$ )





En "*cross-validation*" (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

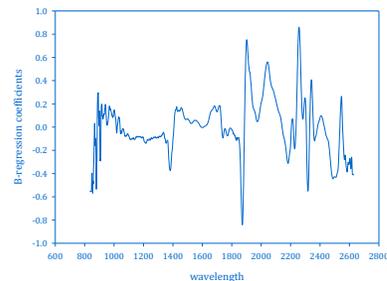
Muestras calibración

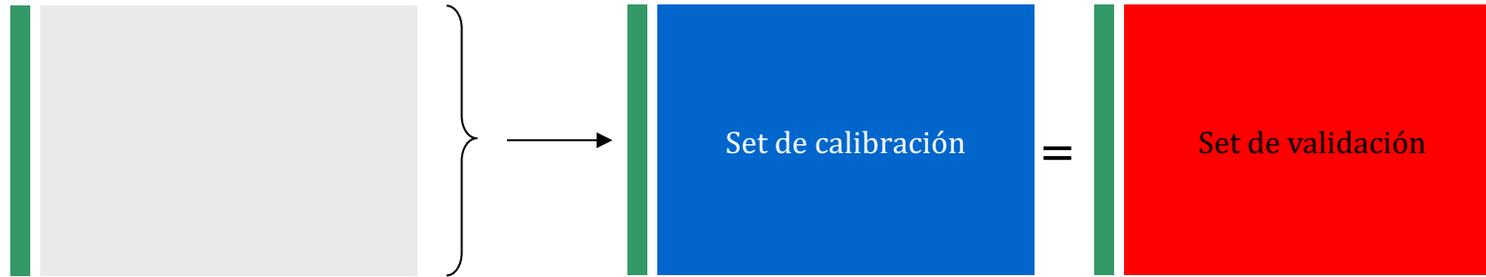


Muestras validación      Estimación  $\hat{y}$



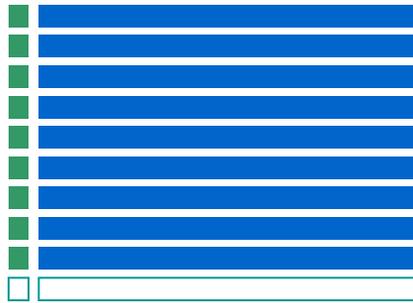
Función de calibración ( $b$ )





En "*cross-validation*" (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

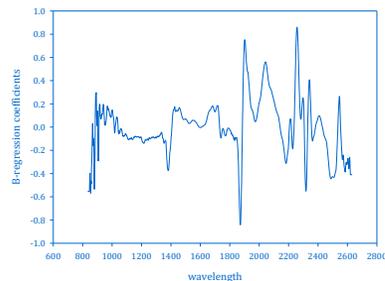
Muestras calibración

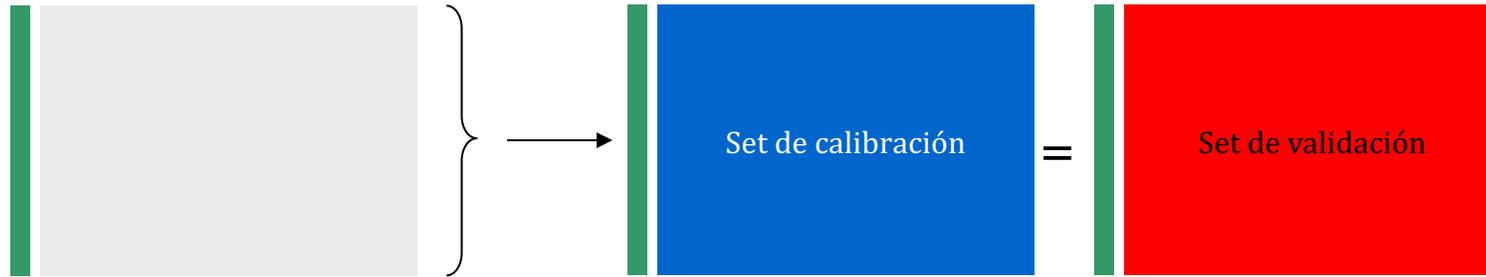


Muestras validación    Estimación  $\hat{y}$



Función de calibración ( $b$ )



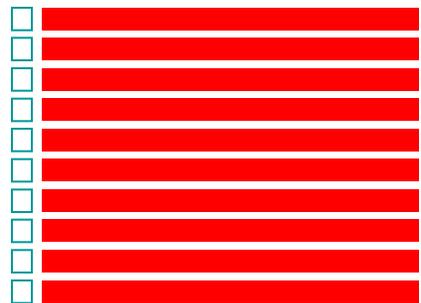


En "*cross-validation*" (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

Muestras calibración



Muestras validación



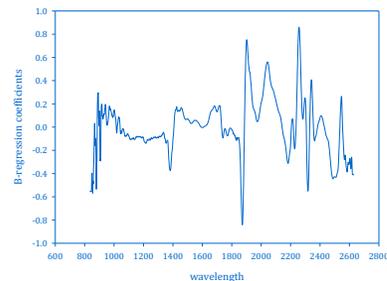
Estimación  $\hat{y}$

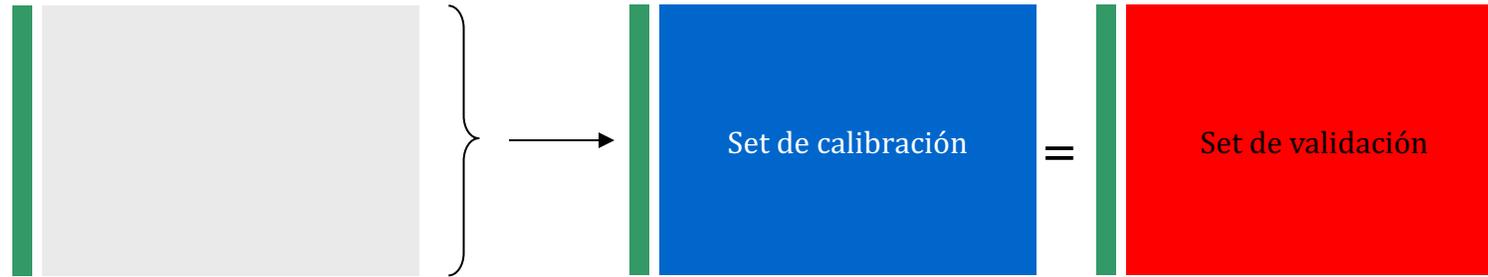


Real  $y$



Función de calibración ( $b$ )



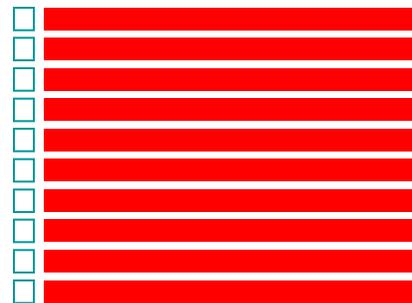


En "*cross-validation*" (validación cruzada) **todas** las muestras sirven para **calibrar** y para **validar**

Muestras calibración



Muestras validación



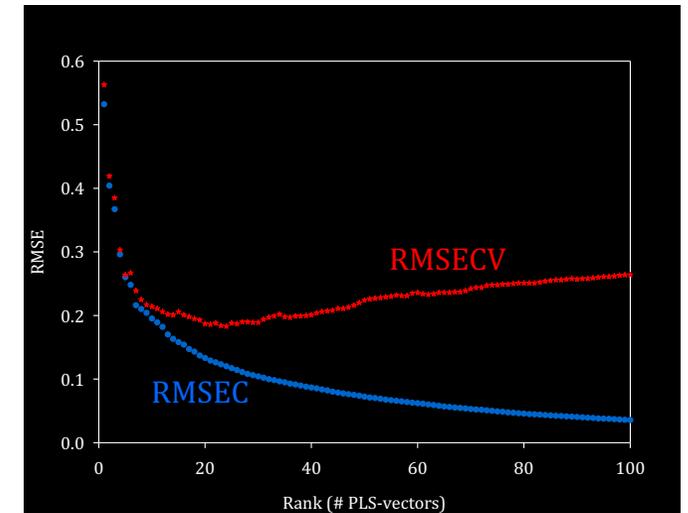
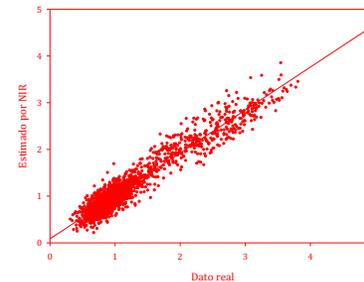
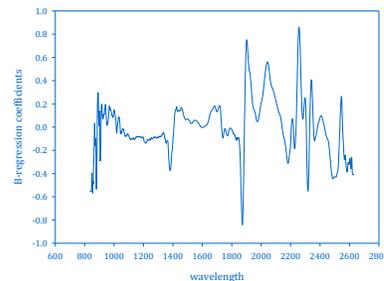
Estimación  $\hat{y}$

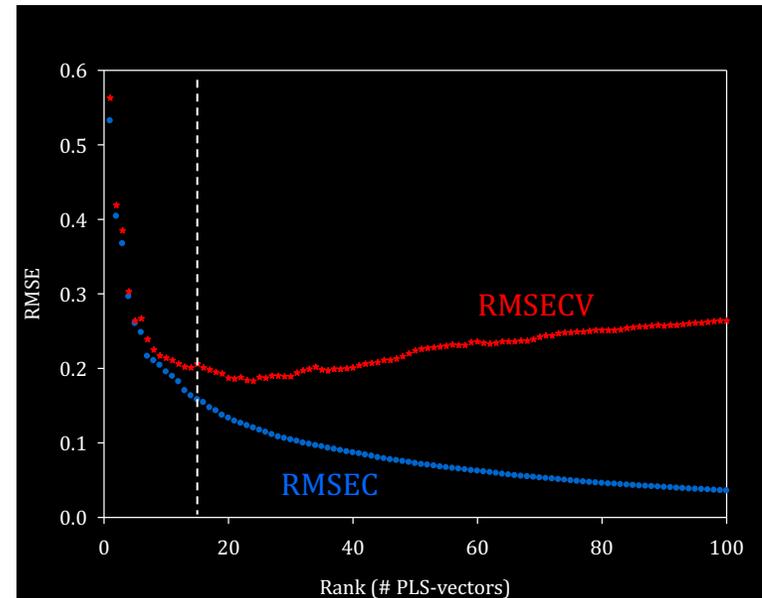
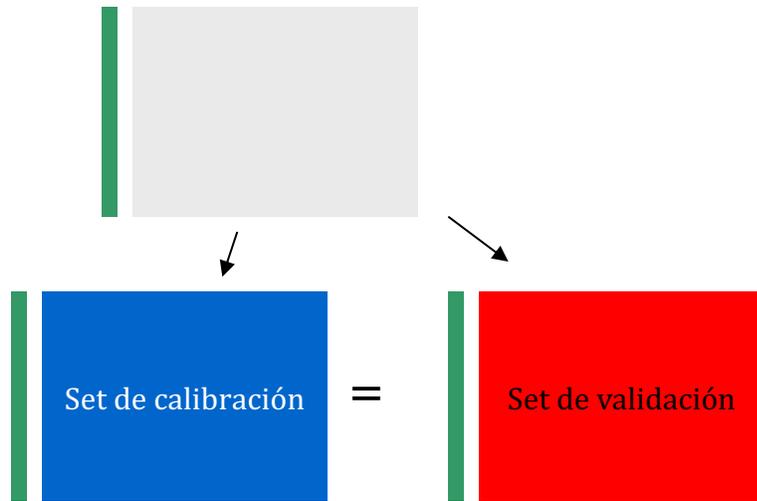


Real  $y$

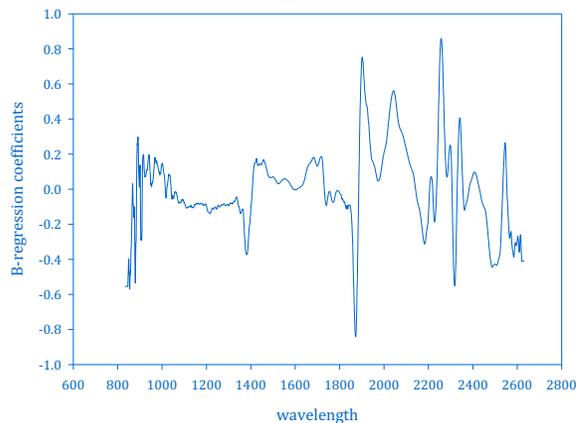


Función de calibración ( $b$ )



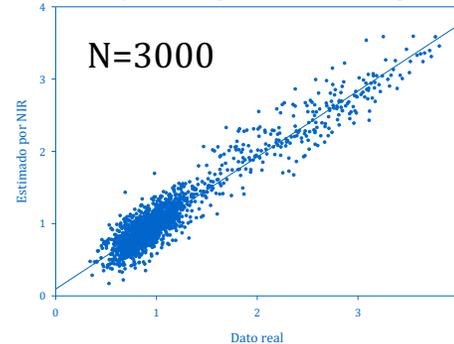


***b*-coefficients**



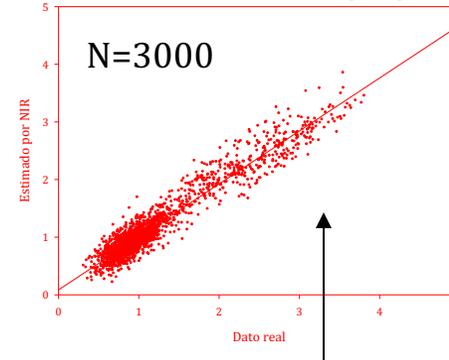
**Función de calibración**

**Ajuste (calibración)**



$R^2$ , RMSEC, *bias*, SEP

**Cross-validación (CV)**



$R^2$ , RMSECV, *bias*, SEP

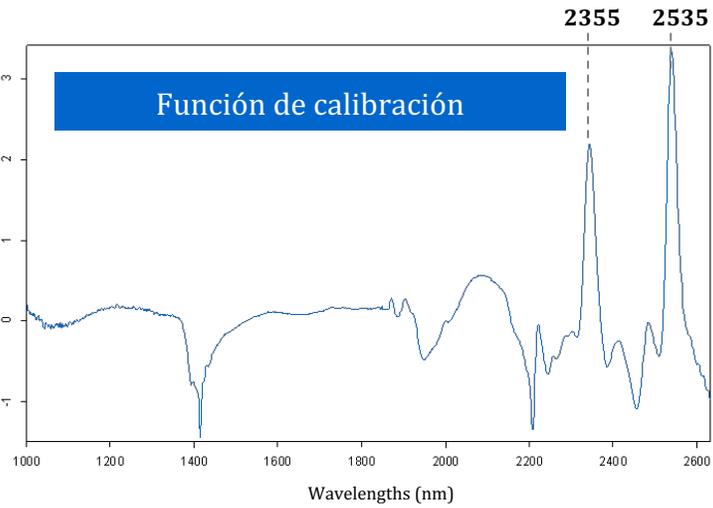
$R^2$ , RMSECV, *bias*, SEP

El parámetro **RMSECV** nos servirá para seleccionar el  $n^{\circ}$  vectores-PLS (ó componentes) óptimo.

No deberíamos llamarle "validación", ya que es ajuste fino.

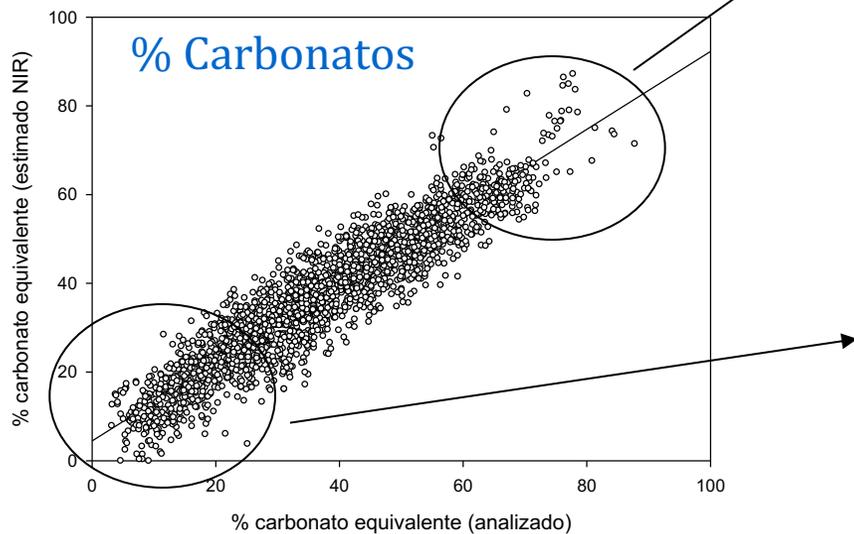
# Coeficientes-*b* regresión: “*el modelo*”

B-regression coefficients



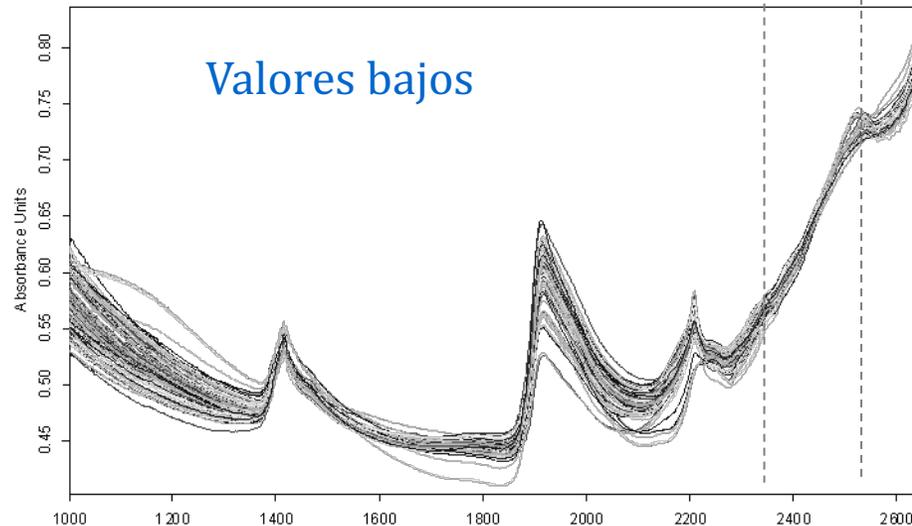
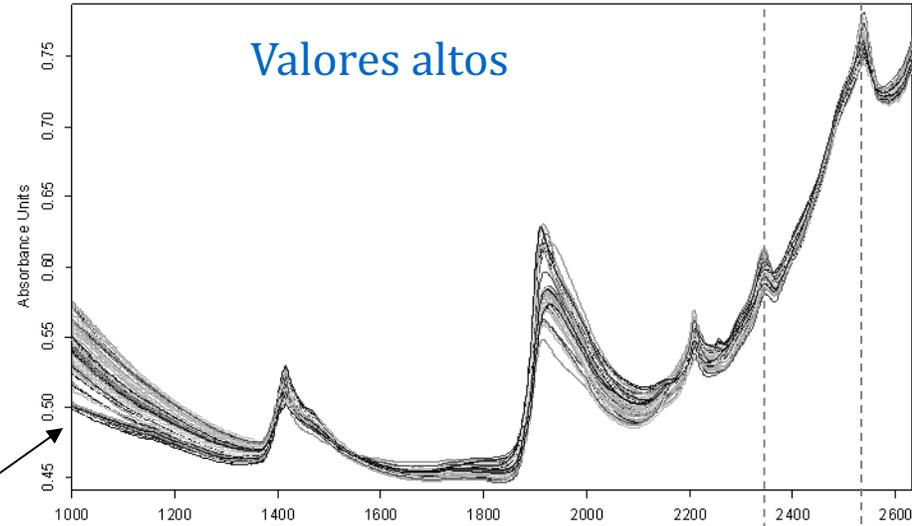
$$y = b_0 + \text{PLSV}_1 \cdot \text{NIR} + \text{PLSV}_2 \cdot \text{NIR} \dots$$

$$y = b_0 + b_1 \cdot \lambda_1 + b_2 \cdot \lambda_2 + b_3 \cdot \lambda_3 \dots + b_p \cdot \lambda_p$$

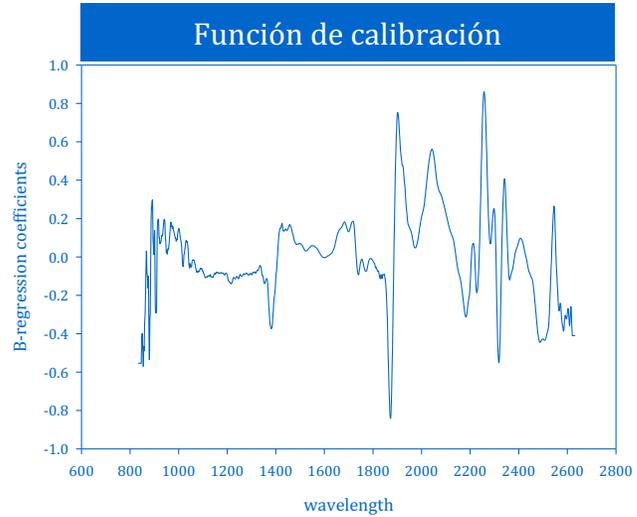


$$\text{Combinación } (2515 \text{ y } 720 \text{ cm}^{-1}) = (2 \cdot 720) + 2515 = 3945 \text{ cm}^{-1} = 2535 \text{ nm}$$

$$2^{\circ}\text{ST } (1415 \text{ cm}^{-1}) = 3 \cdot 1415 = 4245 \text{ cm}^{-1} = 2355 \text{ nm}$$

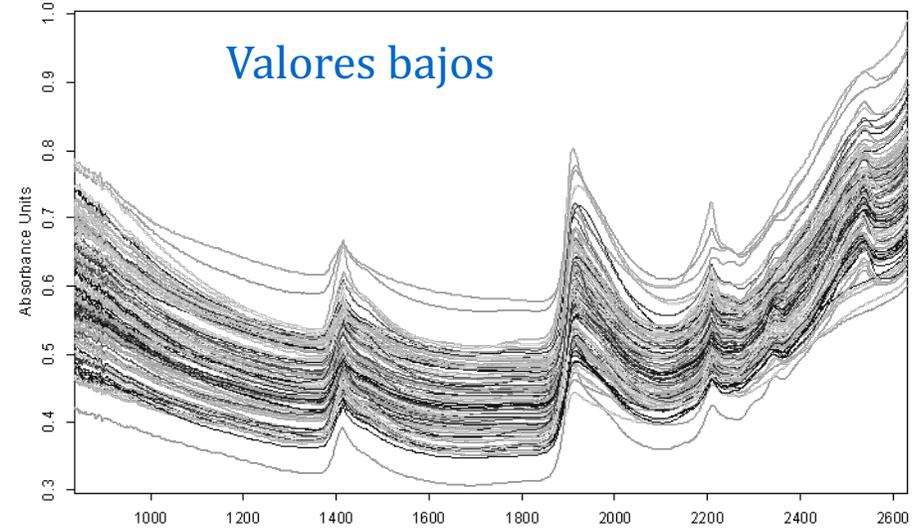
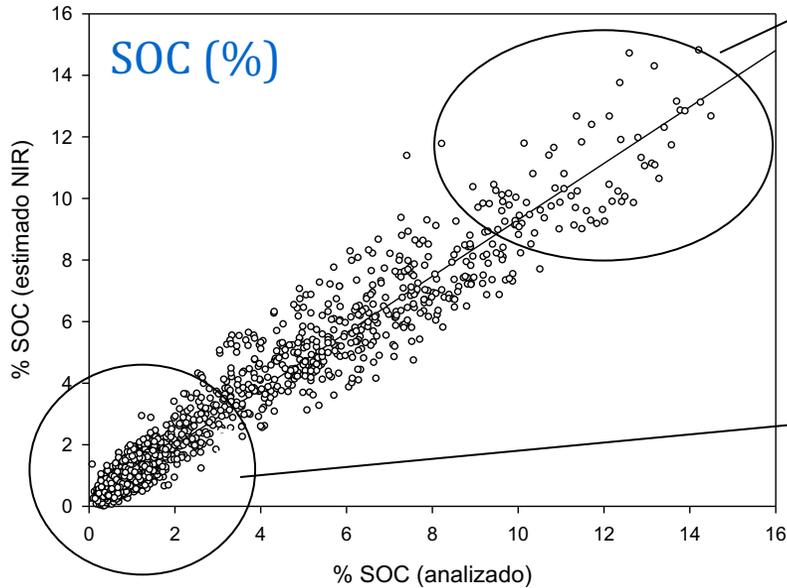
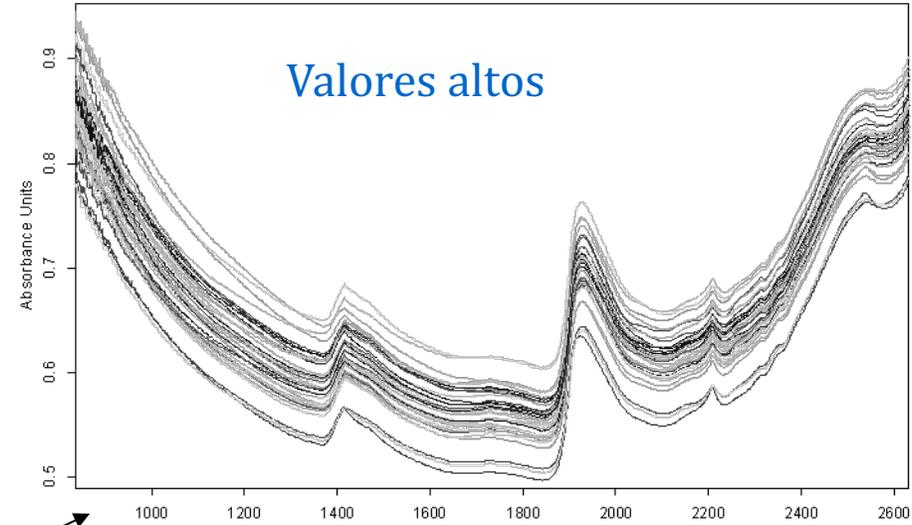


# Coeficientes-*b* regresión: “*el modelo*”



$$y = b_0 + \text{PLSV}_1 \cdot \text{NIR} + \text{PLSV}_2 \cdot \text{NIR} \dots$$

$$y = b_0 + b_1 \cdot \lambda_1 + b_2 \cdot \lambda_2 + b_3 \cdot \lambda_3 \dots + b_p \cdot \lambda_p$$





Muestras calibración



Muestras validación

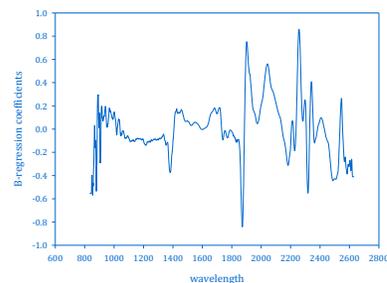


Estimación  $\hat{y}$



*Leave-one-out*

Función de calibración ( $b$ )





Muestras calibración



Muestras validación

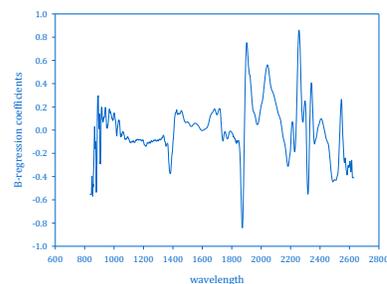


Estimación  $\hat{y}$



*Leave-two-out*

Función de calibración ( $b$ )





Estimación  $\hat{y}$

$k$ -fold cross-validation

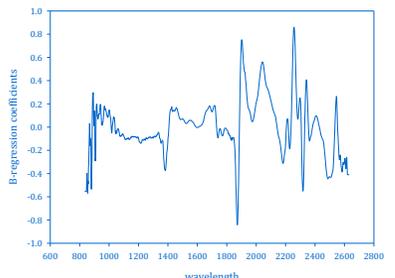
$k$	% total
20	5
10	10
5	20
4	25
2	50

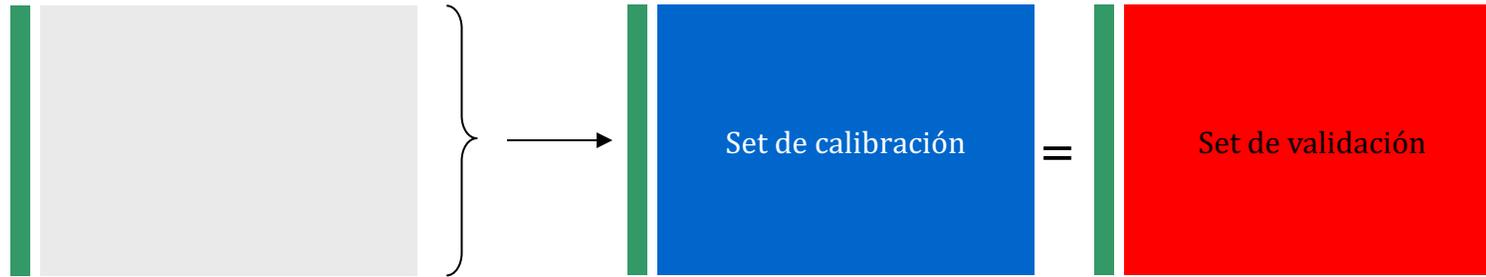
$k$ : número de **segmentos** o particiones

Existen varios métodos para decidir cómo se incluyen las muestras en cada segmento:

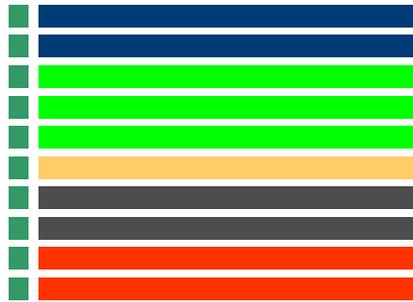
- Bloques consecutivos
- Aleatorio (random)
- Espaciado regular (venetian blinds)
- Manualmente

Función de calibración ( $b$ )





Muestras calibración



Muestras validación



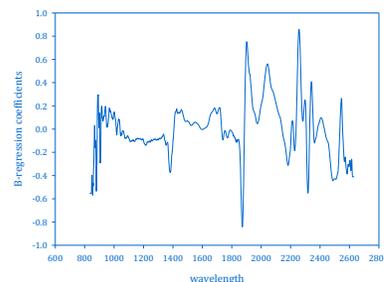
Estimación  $\hat{y}$

*Es importante que las muestras que se sacan y pasan a ser de validación no tengan “réplicas” en la matriz de calibración. Si hay “réplicas”, deben agruparse en un mismo segmento. De otro modo, los resultados serán muy optimistas (**sólo tendremos resultados similares si mantenemos esas condiciones**).*

*Cada uno de esos “grupos” o bloques (mismo segmento) pueden estar formados por muestras de una misma de zona.*

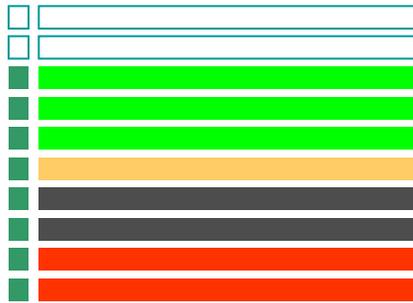
*La validación se debe diseñar de manera que imite las condiciones esperadas durante el uso del modelo en condiciones reales. Por ello, lo que incluya un segmento debe imitar la realidad.*

Función de calibración ( $b$ )

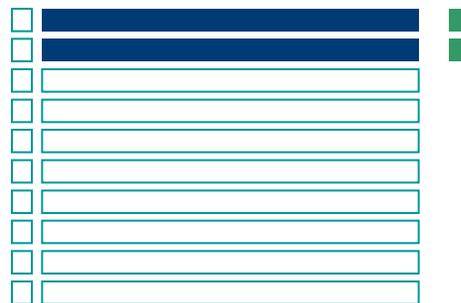




Muestras calibración



Muestras validación



Estimación  $\hat{y}$

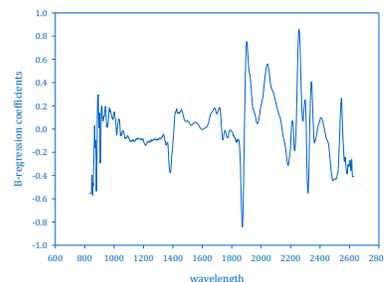


Es importante que las muestras que se sacan y pasan a ser de validación no tengan “réplicas” en la matriz de calibración. Si hay “réplicas”, deben agruparse en un mismo segmento. De otro modo, los resultados serán muy optimistas (**sólo tendremos resultados similares si mantenemos esas condiciones**).

Cada uno de esos “grupos” o bloques (mismo segmento) pueden estar formados por muestras de una misma de zona.

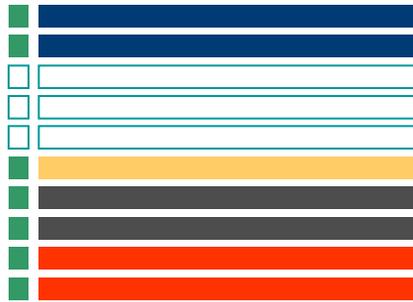
La validación se debe diseñar de manera que imite las condiciones esperadas durante el uso del modelo en condiciones reales. Por ello, lo que incluya un segmento debe imitar la realidad.

Función de calibración ( $b$ )





Muestras calibración



Muestras validación



Estimación  $\hat{y}$

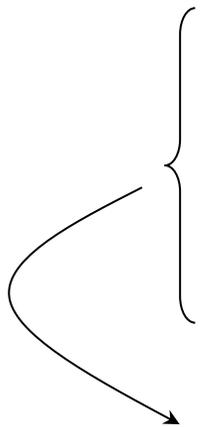
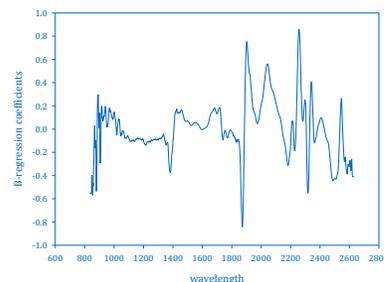


Es importante que las muestras que se sacan y pasan a ser de validación no tengan “réplicas” en la matriz de calibración. Si hay “réplicas”, deben agruparse en un mismo segmento. De otro modo, los resultados serán muy optimistas (**sólo tendremos resultados similares si mantenemos esas condiciones**).

Cada uno de esos “grupos” o bloques (mismo segmento) pueden estar formados por muestras de una misma de zona.

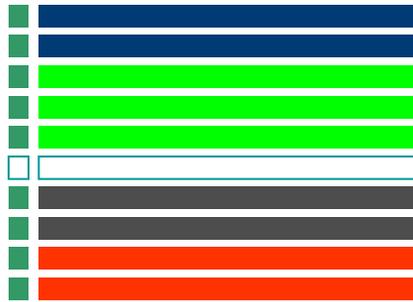
La validación se debe diseñar de manera que imite las condiciones esperadas durante el uso del modelo en condiciones reales. Por ello, lo que incluya un segmento debe imitar la realidad.

Función de calibración ( $b$ )





Muestras calibración



Muestras validación



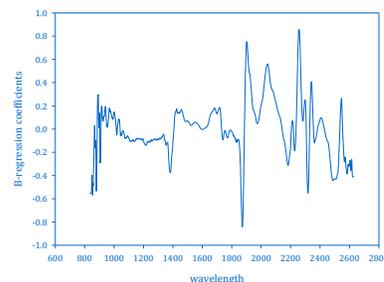
Estimación  $\hat{y}$

Es importante que las muestras que se sacan y pasan a ser de validación no tengan “réplicas” en la matriz de calibración. Si hay “réplicas”, deben agruparse en un mismo segmento. De otro modo, los resultados serán muy optimistas (**sólo tendremos resultados similares si mantenemos esas condiciones**).

Cada uno de esos “grupos” o bloques (mismo segmento) pueden estar formados por muestras de una misma de zona.

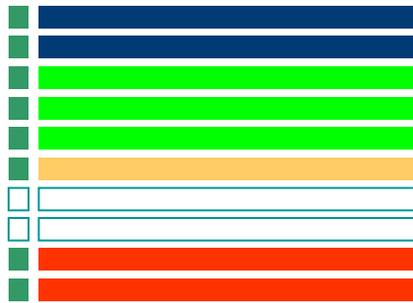
La validación se debe diseñar de manera que imite las condiciones esperadas durante el uso del modelo en condiciones reales. Por ello, lo que incluya un segmento debe imitar la realidad.

Función de calibración ( $b$ )





Muestras calibración



Muestras validación



Estimación  $\hat{y}$

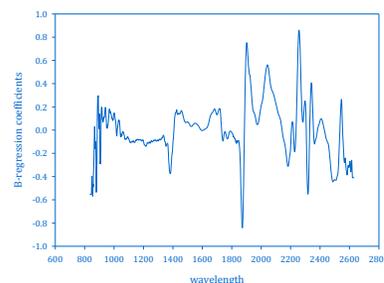


Es importante que las muestras que se sacan y pasan a ser de validación no tengan “réplicas” en la matriz de calibración. Si hay “réplicas”, deben agruparse en un mismo segmento. De otro modo, los resultados serán muy optimistas (**sólo tendremos resultados similares si mantenemos esas condiciones**).

Cada uno de esos “grupos” o bloques (mismo segmento) pueden estar formados por muestras de una misma de zona.

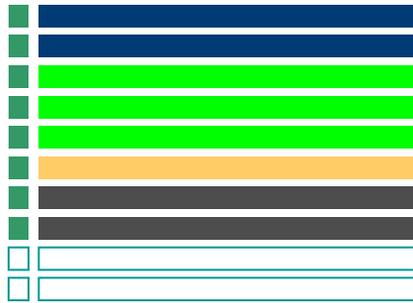
La validación se debe diseñar de manera que imite las condiciones esperadas durante el uso del modelo en condiciones reales. Por ello, lo que incluya un segmento debe imitar la realidad.

Función de calibración ( $b$ )





Muestras calibración



Muestras validación



Estimación  $\hat{y}$

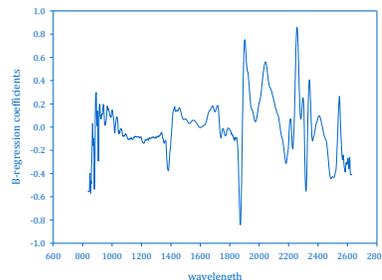


Es importante que las muestras que se sacan y pasan a ser de validación no tengan “réplicas” en la matriz de calibración. Si hay “réplicas”, deben agruparse en un mismo segmento. De otro modo, los resultados serán muy optimistas (**sólo tendremos resultados similares si mantenemos esas condiciones**).

Cada uno de esos “grupos” o bloques (mismo segmento) pueden estar formados por muestras de una misma de zona.

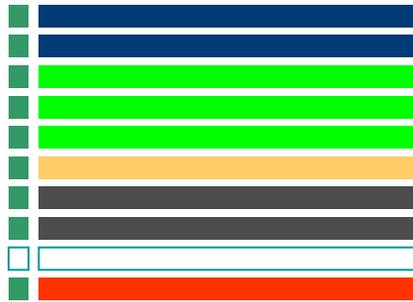
La validación se debe diseñar de manera que imite las condiciones esperadas durante el uso del modelo en condiciones reales. Por ello, lo que incluya un segmento debe imitar la realidad.

Función de calibración ( $b$ )





Muestras calibración



Muestras validación Estimación  $\hat{y}$

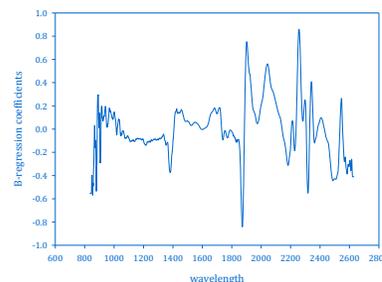


*¿Estas condiciones son realistas? ¿Se van a mantener esas mismas condiciones durante el uso del modelo? ¿voy a usar el modelo de ese mismo modo?*

*Si son condiciones "favorables", pero no realistas, esos resultados serán "aparentemente buenos", pero alejados de la realidad. Las conclusiones sobre la capacidad podrían ser demasiado optimistas, y por ello un tanto erróneas.*

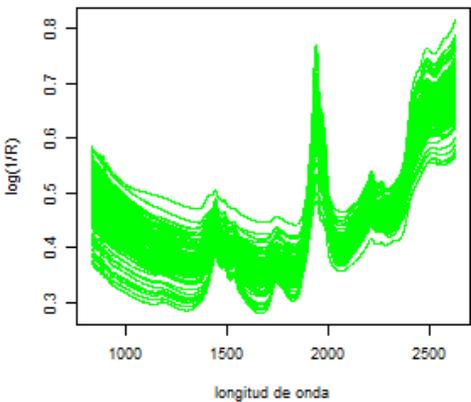
*Sólo obtendré resultados similares si mantengo las mismas condiciones.*

Función de calibración ( $b$ )

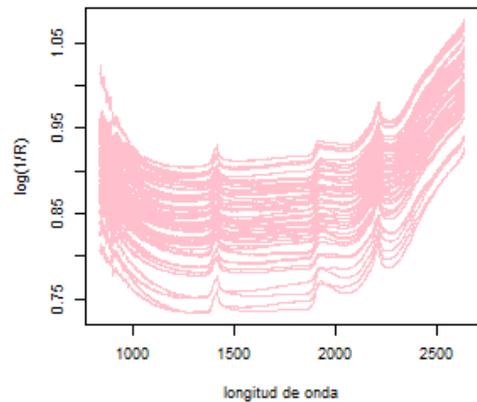


## Suelos de 10 zonas

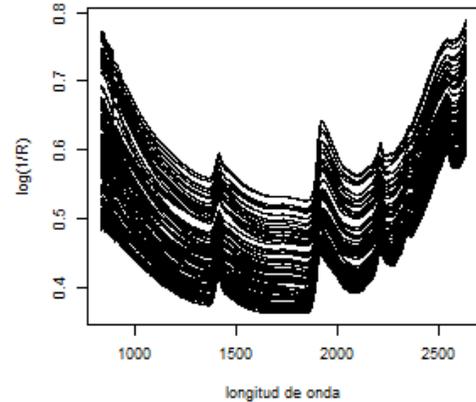
sitio 1



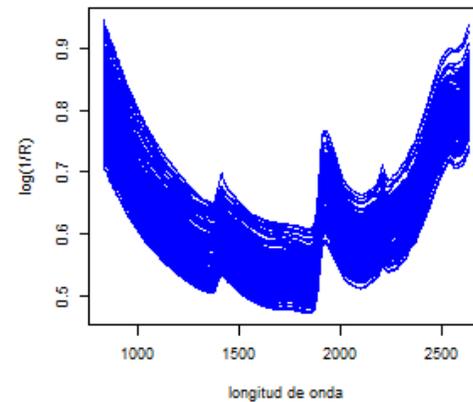
sitio 2



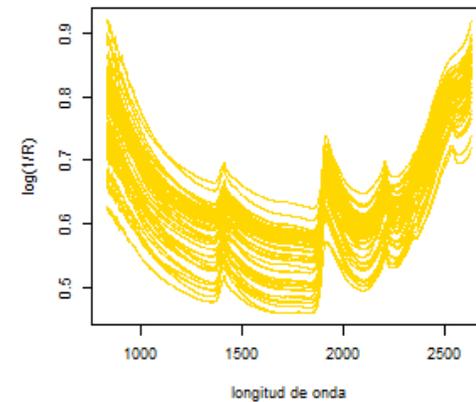
sitio 3



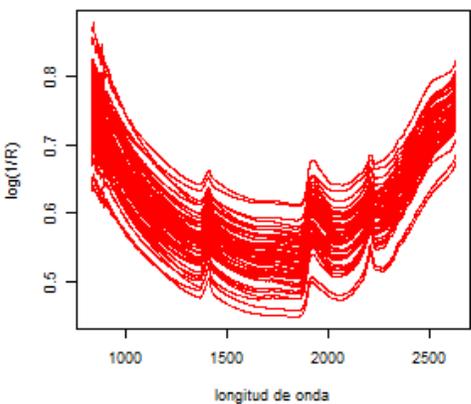
sitio 4



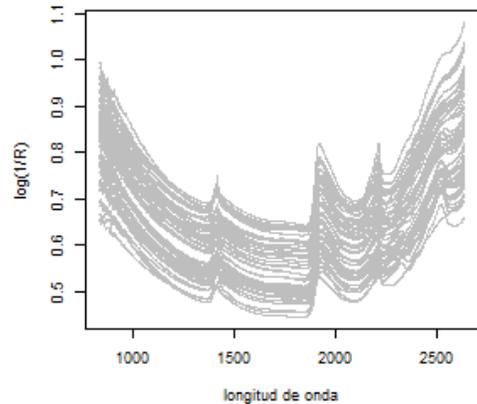
sitio 5



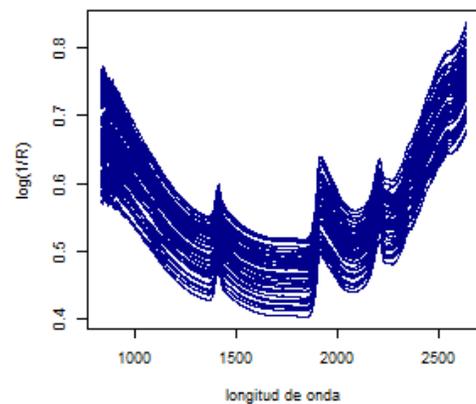
sitio 6



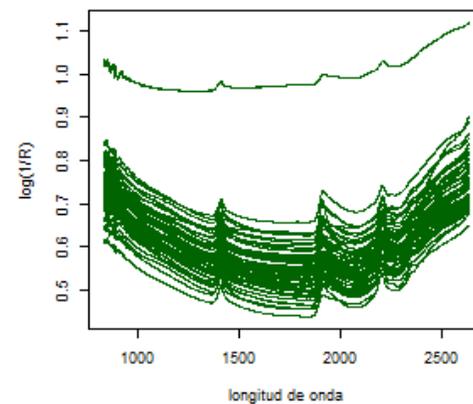
sitio 7



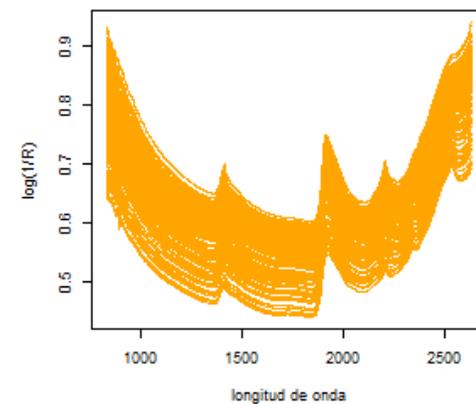
sitio 8



sitio 9



sitio 10



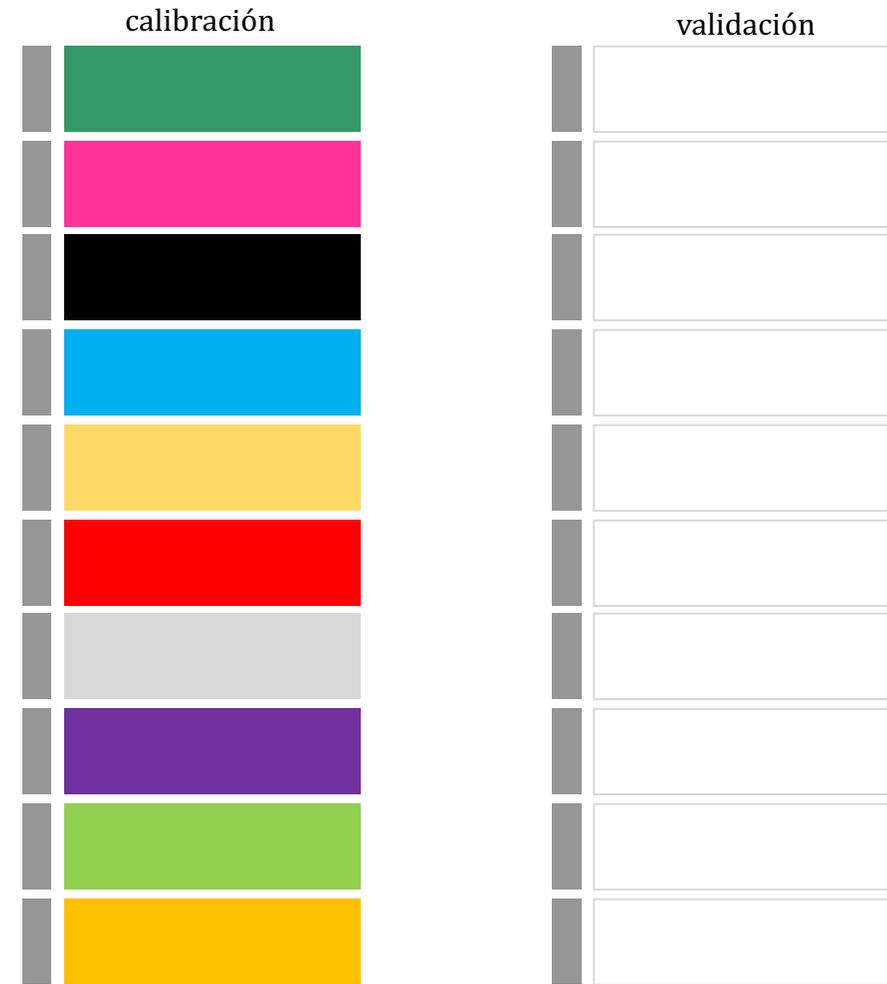
Suelos de 10 zonas:

*Cross-Validation* **random** (10 segmentos "random"): las muestras que forman parte de un segmento son seleccionadas al azar.



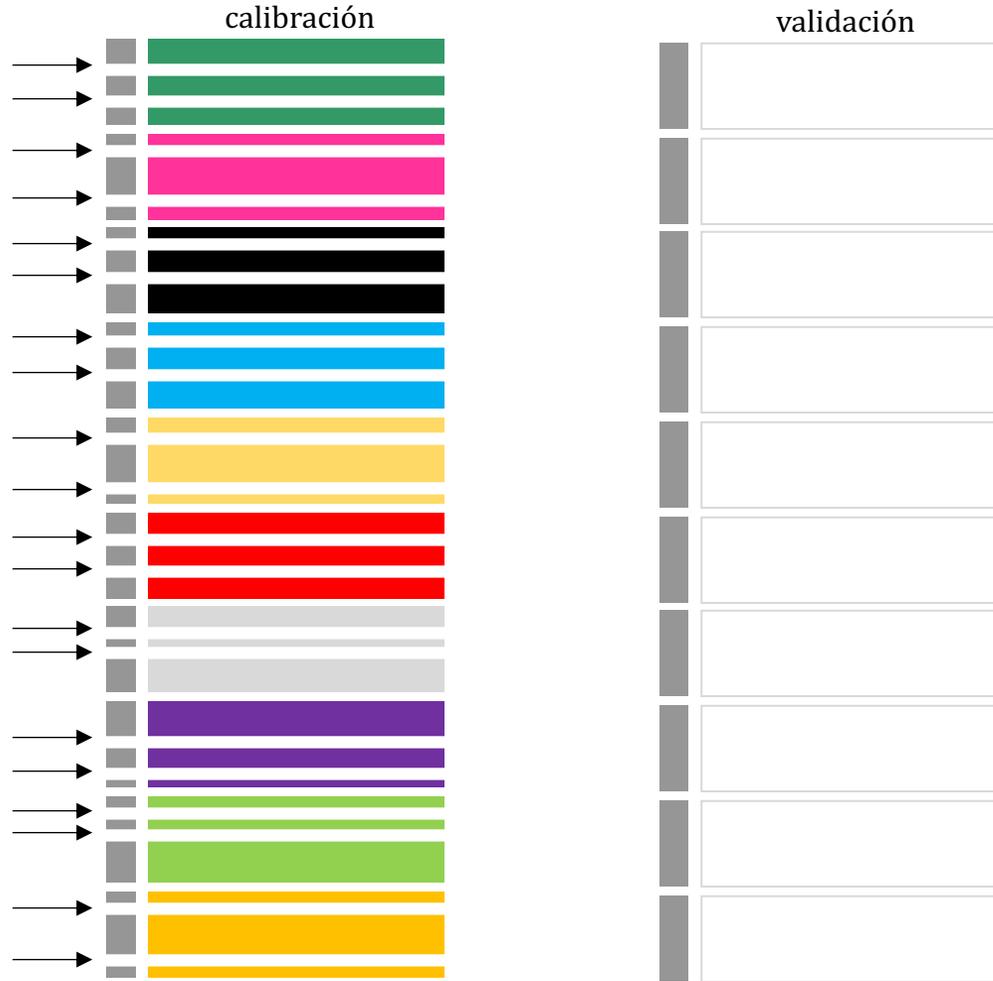
Suelos de 10 zonas:

*Cross-Validation* "**por sitio**" (segmentos "sitio"): todas las muestras de un mismo sitio son medidas en un mismo segmento.



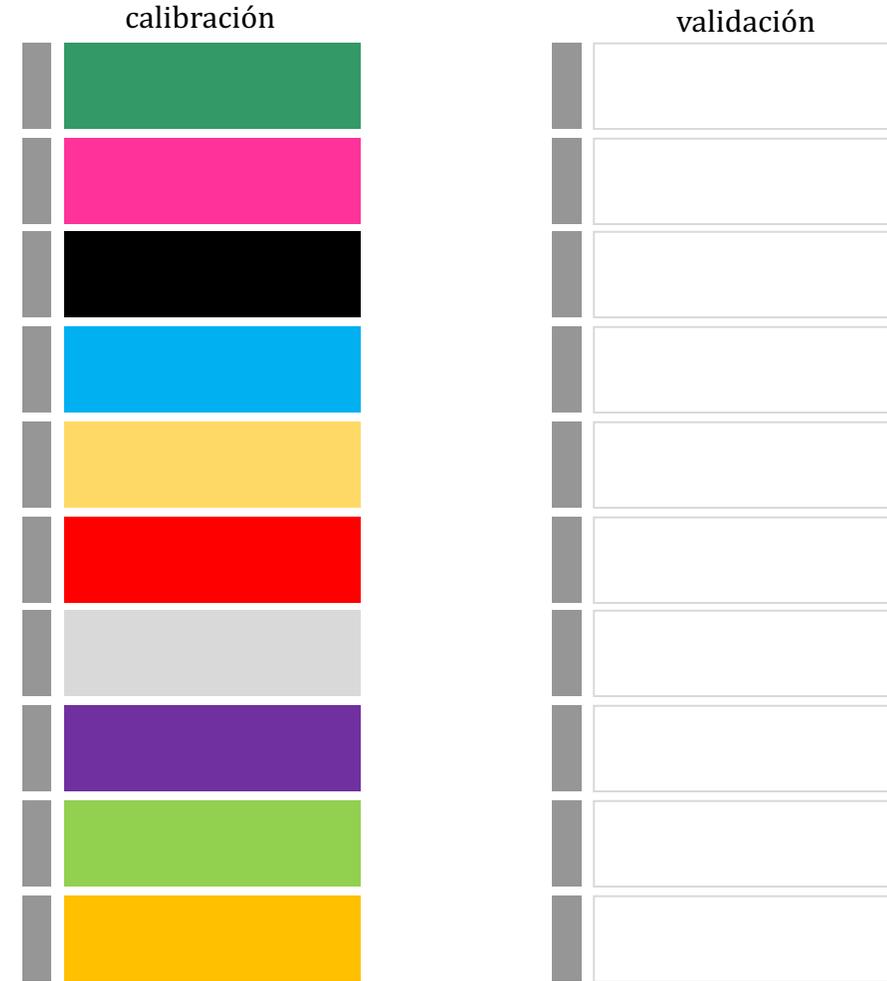
Suelos de 10 zonas:

*Cross-Validation* **random** (10 segmentos "random"): las muestras que forman parte de un segmento son seleccionadas al azar.



Suelos de 10 zonas:

*Cross-Validation* "**por sitio**" (segmentos "sitio"): todas las muestras de un mismo sitio son medidas en un mismo segmento.



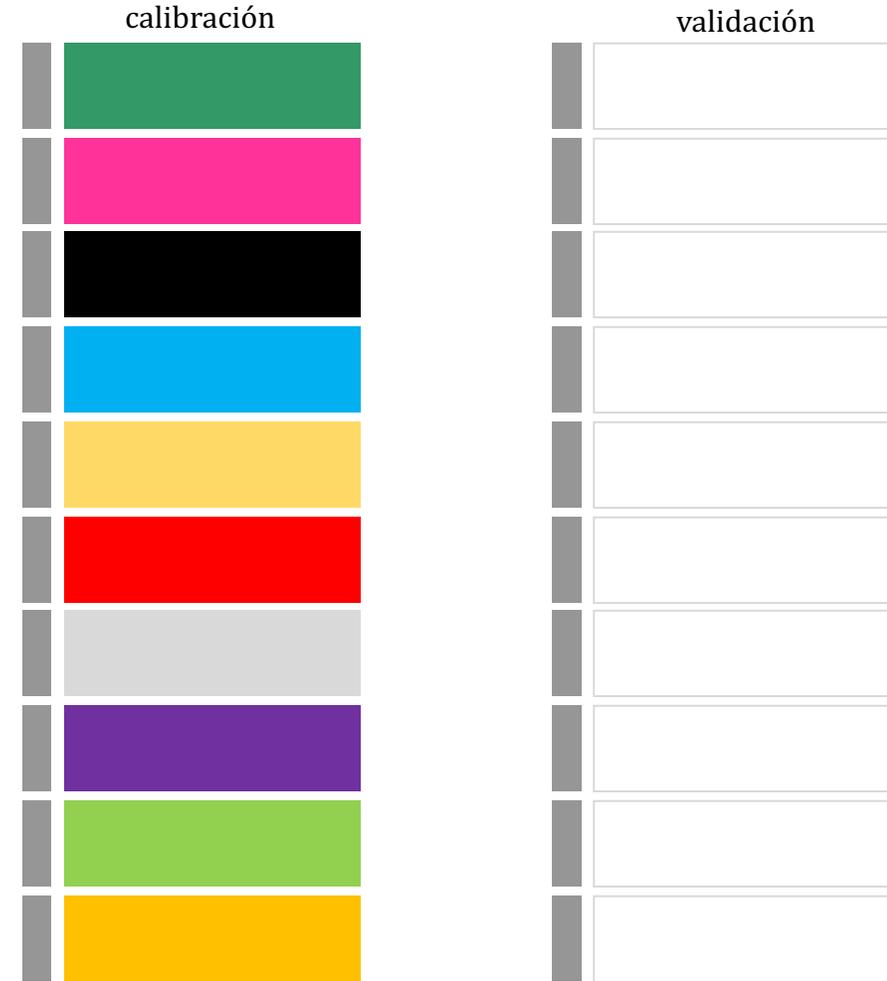
Suelos de 10 zonas:

*Cross-Validation* **random** (10 segmentos "random"): las muestras que forman parte de un segmento son seleccionadas al azar.



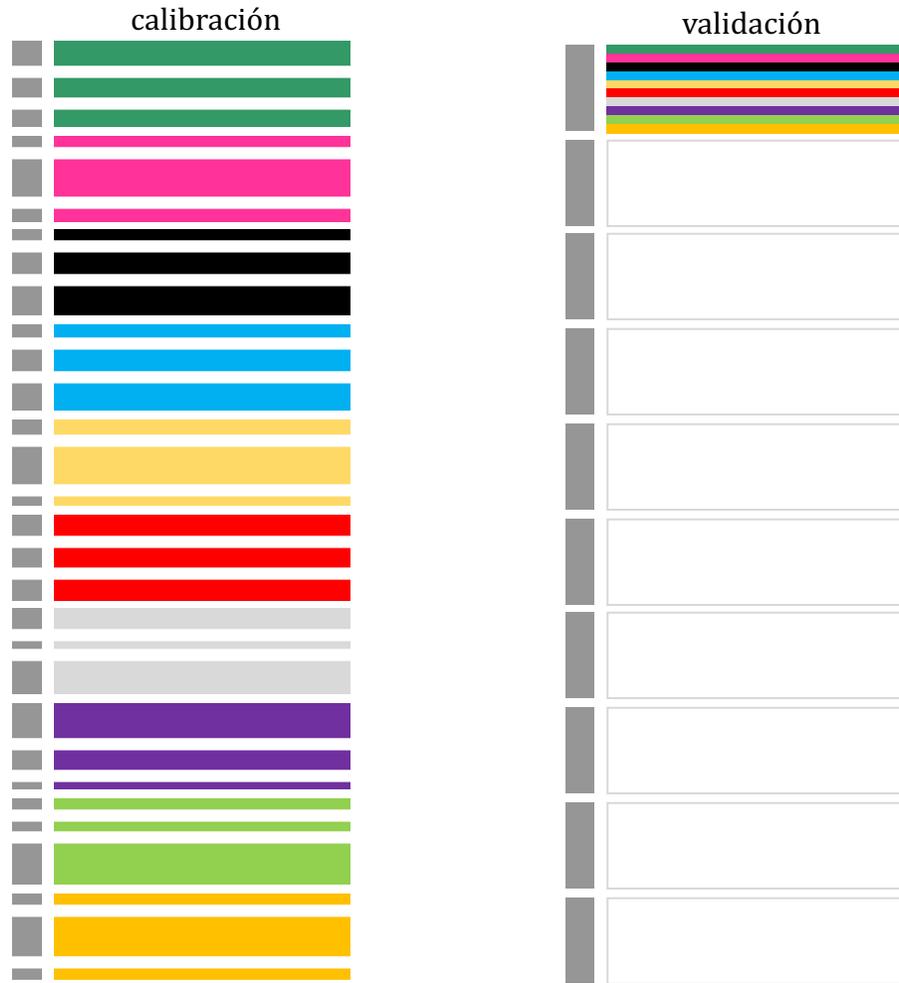
Suelos de 10 zonas:

*Cross-Validation* "**por sitio**" (segmentos "sitio"): todas las muestras de un mismo sitio son medidas en un mismo segmento.



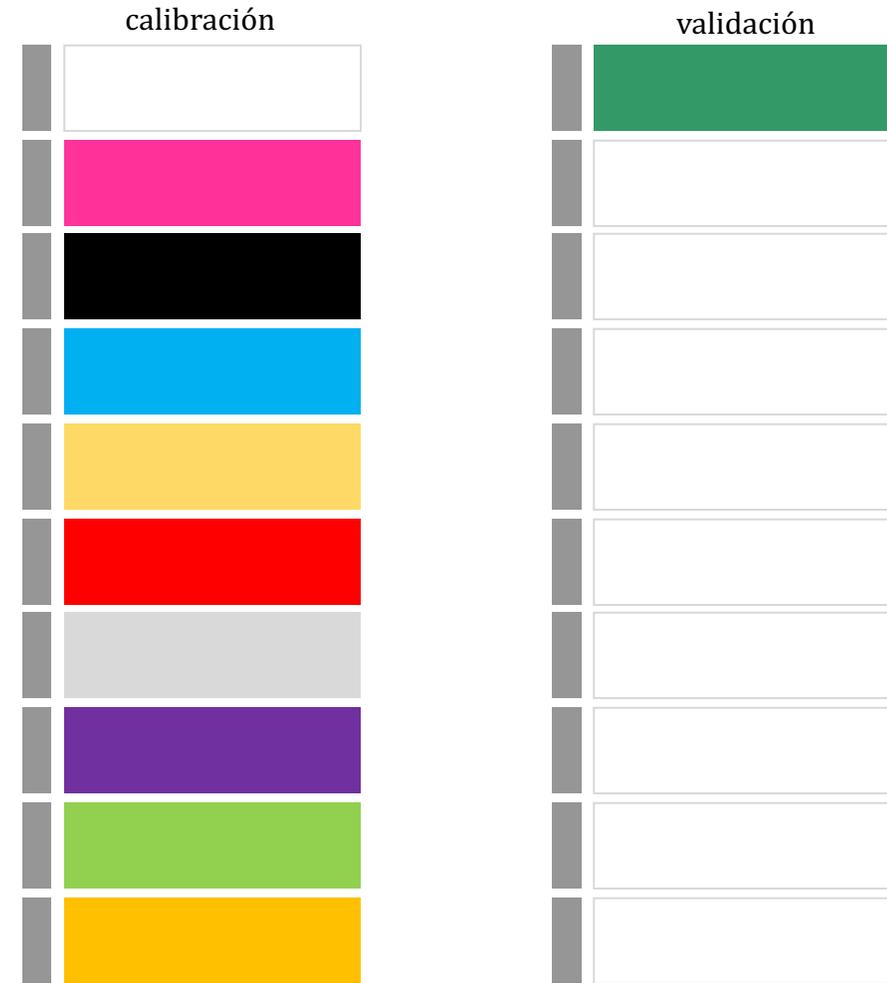
Suelos de 10 zonas:

*Cross-Validation* **random** (10 segmentos "random"): las muestras que forman parte de un segmento son seleccionadas al azar.



Suelos de 10 zonas:

*Cross-Validation* "**por sitio**" (segmentos "sitio"): todas las muestras de un mismo sitio son medidas en un mismo segmento.



Suelos de 10 zonas:

*Cross-Validation* **random** (10 segmentos "random"): las muestras que forman parte de un segmento son seleccionadas al azar.



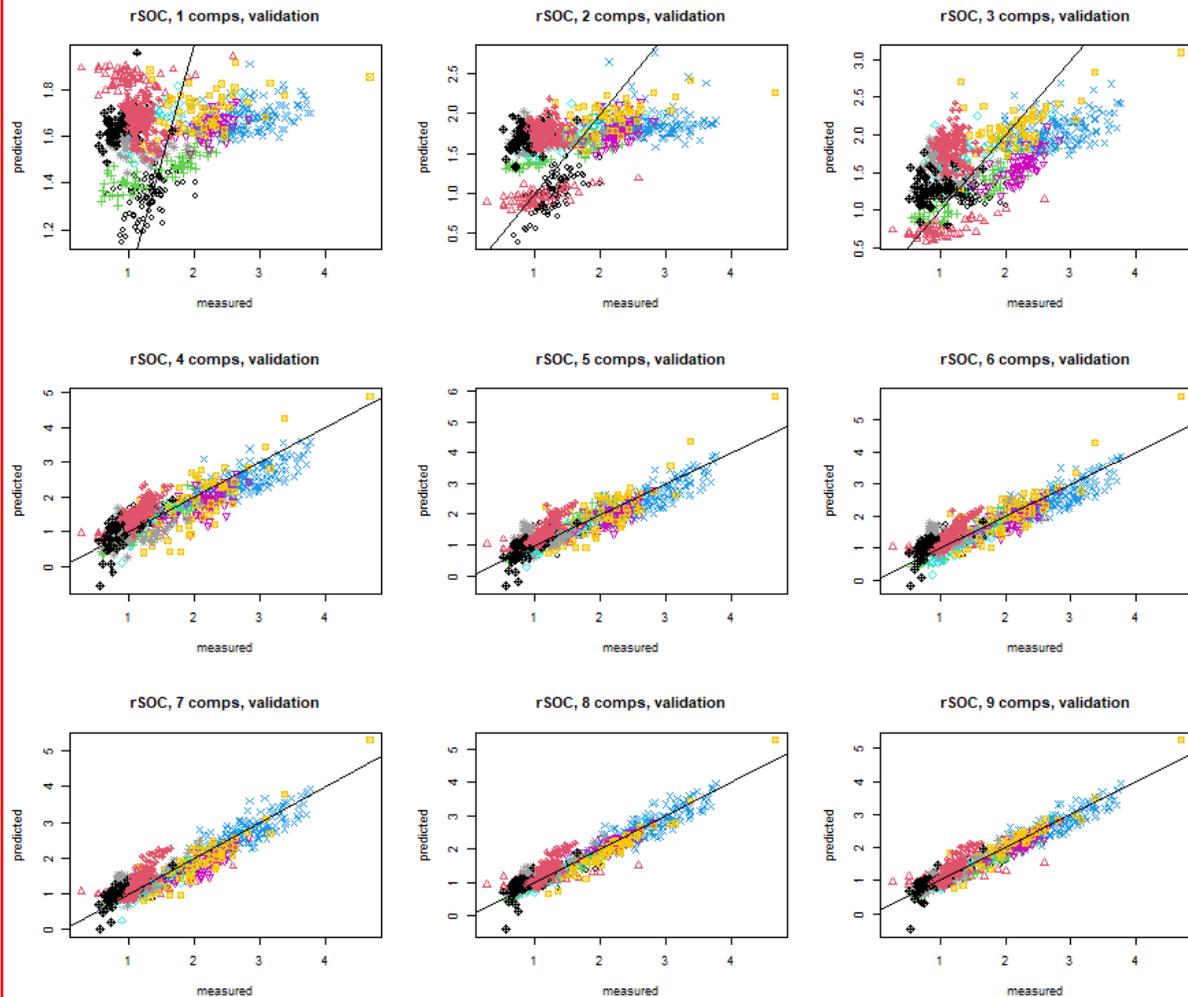
Suelos de 10 zonas:

*Cross-Validation* "**por sitio**" (segmentos "sitio"): todas las muestras de un mismo sitio son medidas en un mismo segmento.



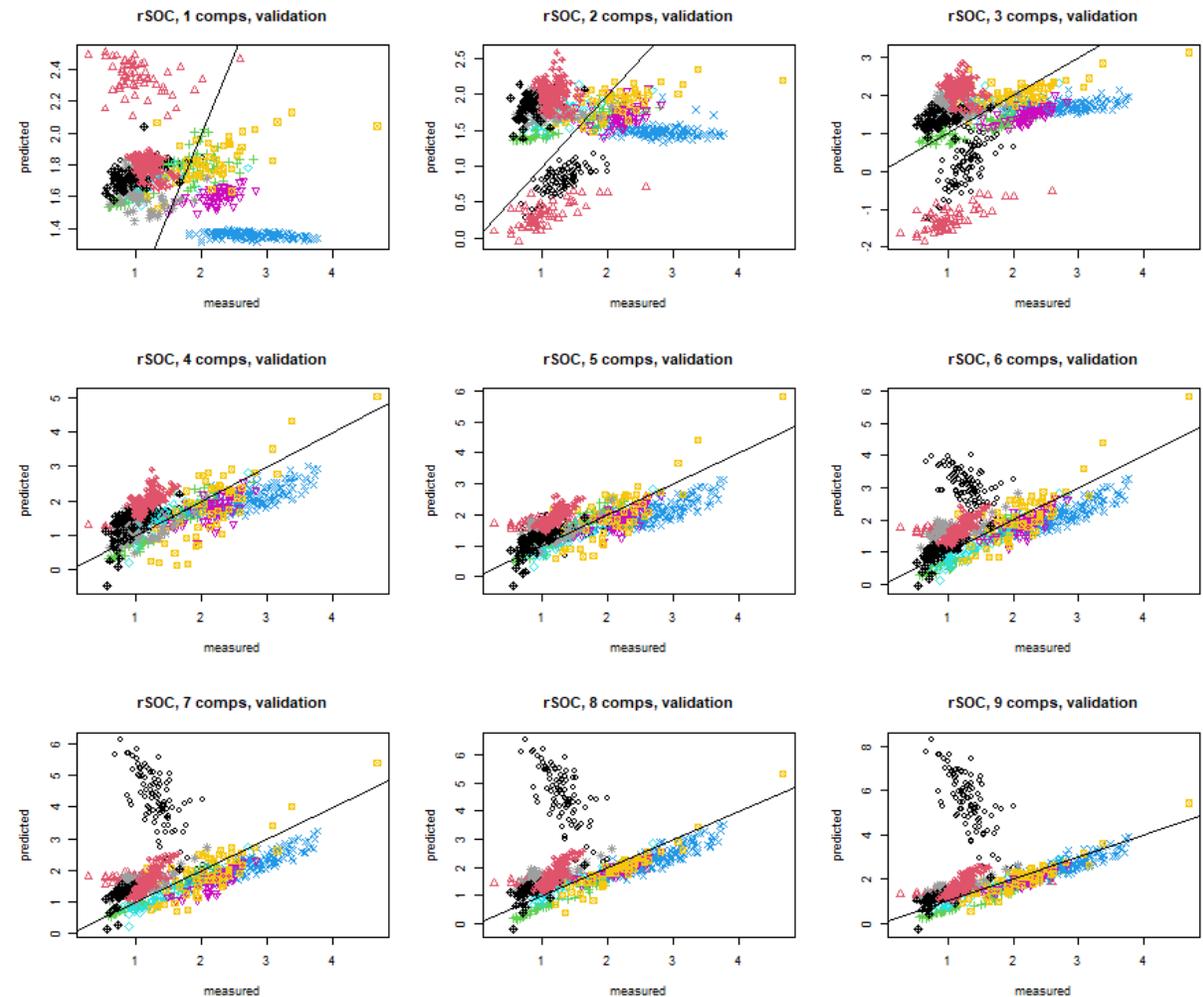
Suelos de 10 zonas:

*Cross-Validation* **random** (10 segmentos "random"): las muestras que forman parte de un segmento son seleccionadas al azar.



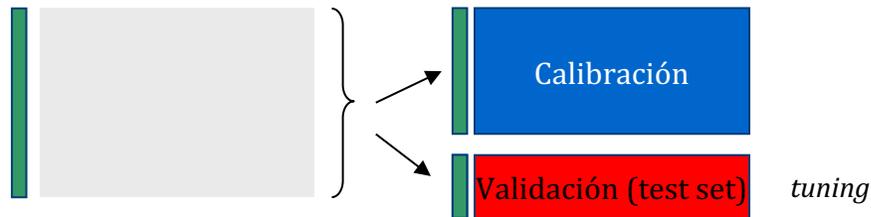
Suelos de 10 zonas:

*Cross-Validation* "**por sitio**" (segmentos "sitio"): todas las muestras de un mismo sitio son medidas en un mismo segmento.

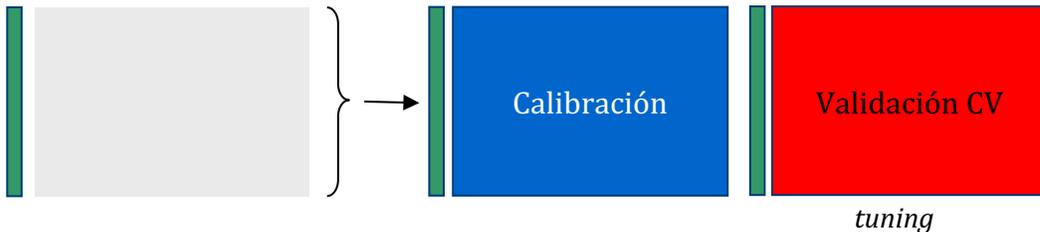


## Training

### Validación *test set*

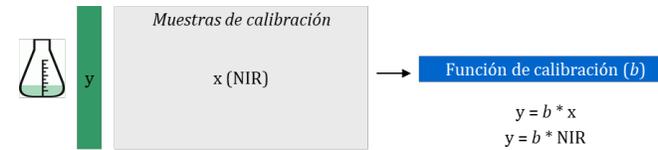


### Validación *cross-validation*

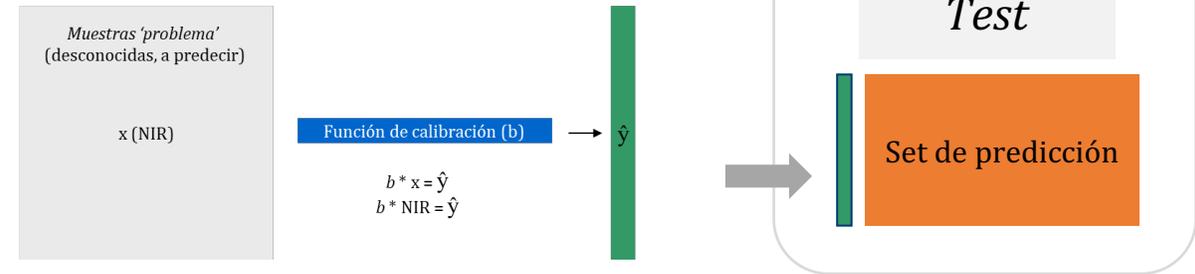


- A las **validaciones** que son **usadas para decidir** el **número de componentes** o **vectores-PLS** (rango del modelo) y para seleccionar cuál es el **pre-procesamiento** más adecuado **no se les debería de identificar como "validaciones"**, ya que estamos **terminando de definir** la arquitectura y parámetros del modelo ("*fine tuning*"= ajuste fino). No es una verdadera "validación".
- Todo esto es **entrenamiento** ("*training*") -> obtener el modelo

### 1º) Desarrollo de la calibración (o modelo)



### 2º) Uso de la calibración



- **Una vez el modelo está totalmente definido**, lo ideal es proceder a **probarlo** en un nuevo grupo de muestras (**set de predicción**), **simulando su uso real**, para ver **cómo funciona** (con qué exactitud lo hace). Esto **sí es una verdadera "validación"** (sentido estricto). Algunos autores -> "**validación externa**"
- El set de predicción debe estar formado por muestras independientes de las usadas anteriormente, pero dentro de los **límites** del modelo. **La selección de muestras incluidas debe simular un uso real (evitar situaciones poco realistas!)**.
- Los **resultados** obtenidos (evaluados con  $R^2$ , RMSEP, bias y SEP) son los que deberíamos mostrar para **reportar la capacidad** del modelo, y **nunca** para re-ajustar los parámetros del modelo ( $n^\circ$  de componentes, pre-procesamiento, etc.).

\*  $R^2$  (coeficiente de determinación):

- es el parámetro más usado, pero probablemente sea el peor para evaluar la calidad de las predicciones.
- Sólo informa de la relación lineal entre el dato real y el estimado.
- Su valor es dependiente del rango de valores, y está afectado por valores “extremos”.

\* RMSEP (*root mean square error of prediction*):

- es una medida de la exactitud de la predicción.
- Error total esperado (expresado en mismas unidades)

$$RMSEP = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (Differ_i)^2} \quad Differ_i = \hat{y}_i \text{ estimado} - y_i \text{ real} = \text{residual}$$

\* *Bias*:

- Error sistemático. Cálculo sencillo: diferencia entre la media de los valores estimados y la media de los valores reales.

\* SEP (*standard error of prediction*):

- Es una medida de la precisión de la predicción (desviación estándar de los residuales). Es la parte del error que no se debe al *bias*.

$$SEP = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (Differ_i - bias)^2} \quad RMSEP^2 \approx bias^2 + SEP^2$$

\* RPD (*residual predictive deviation, ratio of prediction to deviation, ratio performance deviation*)

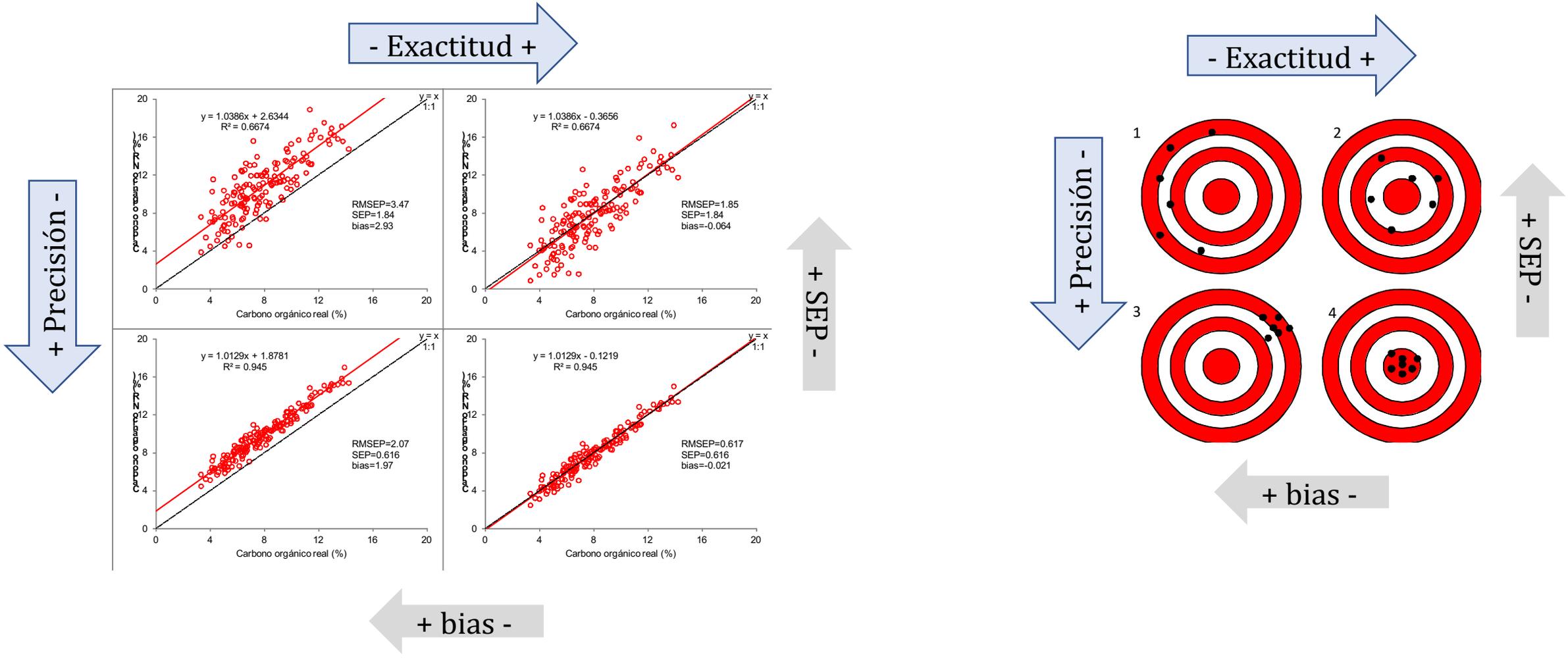
Para el cálculo de la RPD es necesario calcular la desviación estándar (SD) de los valores reales de las muestras

$$RPD = (SD/SEP) \text{ ó } RPD = (SD/RMSEP)$$

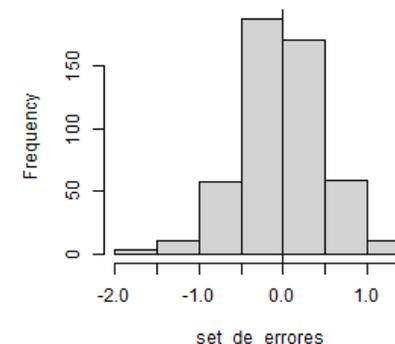
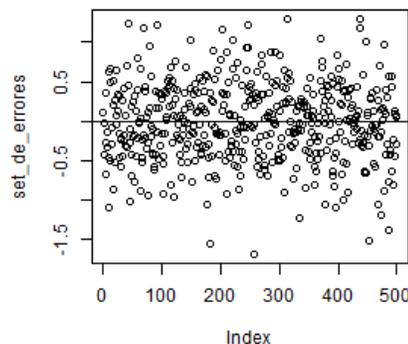
Parámetro muy usado, pero también dependiente del rango de valores de “y”. Útil para comparar calibraciones (sólo si cubren rangos similares).  $RPD > 2$

\* RPIQ (parecido a RPD, pero usando el rango intercuartílico en lugar de SD, con objeto de atenuar valores extremos)

$$RMSEP^2 \approx bias^2 + SEP^2$$



# Generamos aleatoriamente un set de 500 errores, con distribución normal, con media 0 y desviación 0.5:  
`> set_de_errores <- rnorm(500, 0, 0.5)`



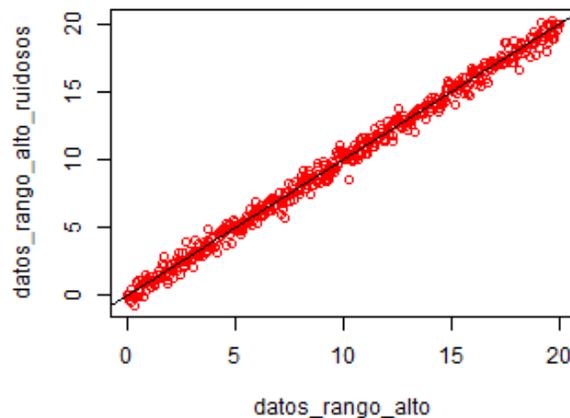
# Generamos un set de 500 muestras, simulando que es un set con **rango alto** (valores desde **0** hasta **20**)

`> datos_rango_alto <- seq(from=0, to=20, length=500)`

# Ahora **añadimos** el set de **errores** para **simular "mediciones ruidosas"**

`> datos_rango_alto_ruidosos <- datos_rango_alto + set_de_errores`

**Rango alto**



**Rango alto**  
 $R^2=0.993$   
 RMSEP=0.5

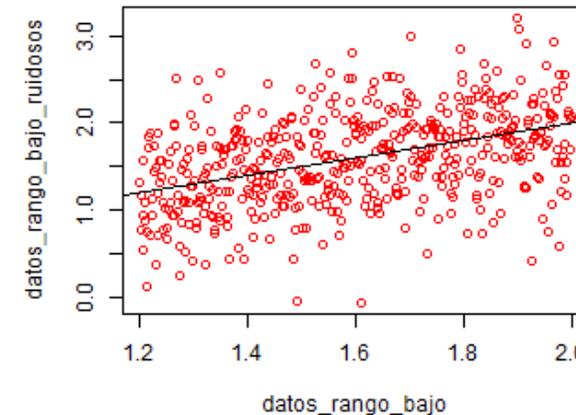
# Ahora generamos un set de 500 muestras, simulando que es un set con **rango bajo** (valores desde **1.2** hasta **2**)

`> datos_rango_bajo <- seq(from=1.20, to=2, length=500)`

# Ahora **añadimos** el **mismo** set de **errores** para **simular "mediciones ruidosas"**

`> datos_rango_bajo_ruidosos <- datos_rango_bajo + set_de_errores`

**Rango bajo**



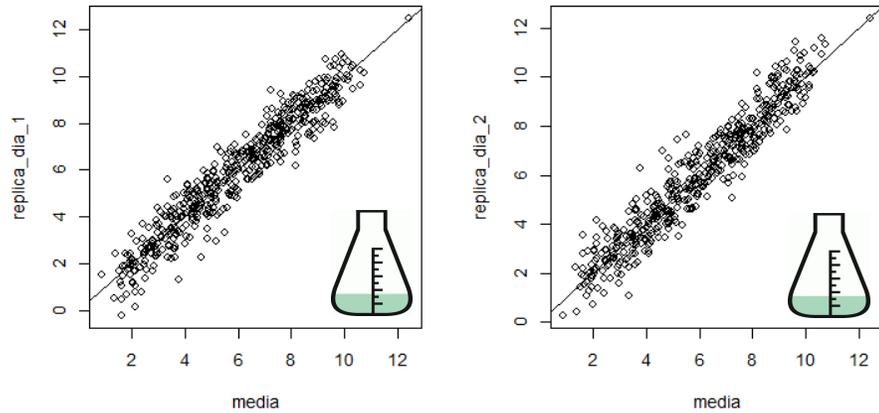
**Rango bajo**  
 $R^2=0.169$   
 RMSEP=0.5

¿?

**¡Mismo error!**  
**¿Mismas conclusiones?**

## Reproducibilidad del método de referencia

Mismas muestras analizadas en días diferentes

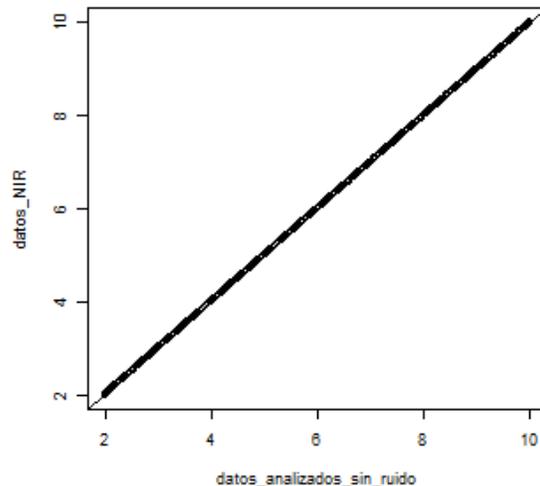


\* **Parte del error que se cuantifica procede del método de referencia:**

- Puedes reducirlo haciendo repeticiones (réplicas analíticas) del set de muestras de predicción
- Sin embargo, hacer esas repeticiones para las muestras de calibración podría no ser tan necesario, y ese esfuerzo dedicarlo a analizar otra muestra que aumente la representatividad del modelo.

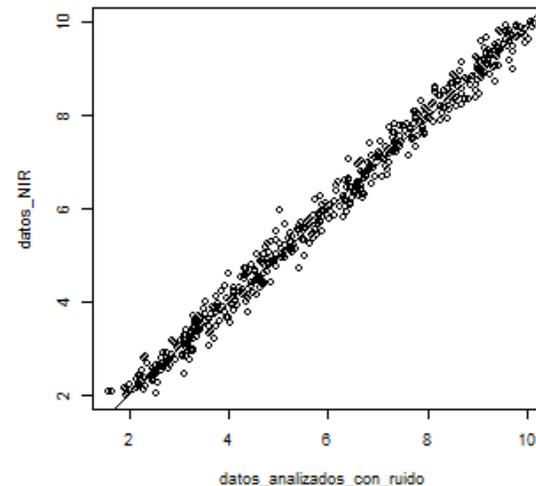
### Supuesto 1º

Estimación NIR perfecta vs método referencia perfecto



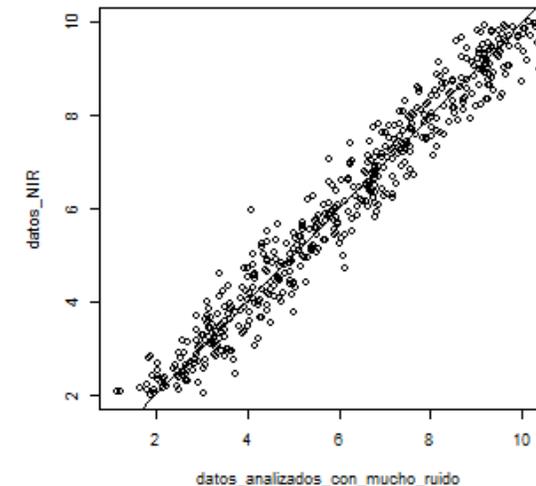
### Supuesto 2º

Estimación NIR perfecta vs método referencia poco ruido



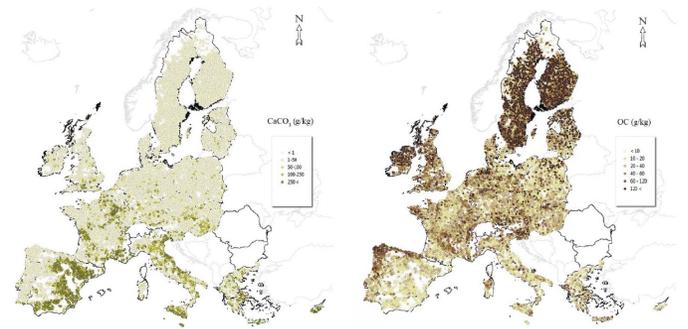
### Supuesto 3º

Estimación NIR perfecta vs método referencia mucho ruido



## LUCAS: Land Use/Cover Area frame Statistical Survey

- \* Espectros vis-NIR (400-2500 nm)
- \* Varias propiedades edáficas:
  - pH (H<sub>2</sub>O)
  - pH (CaCl<sub>2</sub>)
  - SOC
  - N
  - Carbonatos
  - Textura...

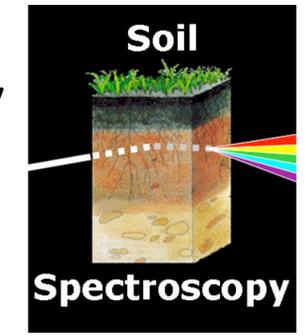


~20.000 muestras

Accesible <http://eusoils.jrc.ec.europa.eu/projects/Lucas/>

## The Soil Spectroscopy Group

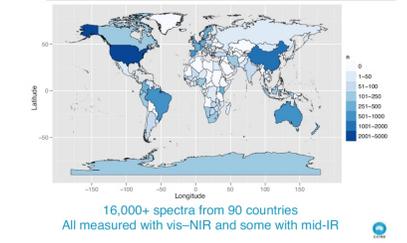
### Making a Global Soil Spectral Library



## Raphael VISCARRA ROSSEL

Soil & landscape program CSIRO Land & Water

### The global soil spectral library - GSSL



Earth-Science Reviews 155 (2016) 198–230

Contents lists available at ScienceDirect

Earth-Science Reviews

journal homepage: [www.elsevier.com/locate/earscirev](http://www.elsevier.com/locate/earscirev)

A global spectral library to characterize the world's soil

R.A. Viscarra Rossel<sup>a,\*</sup>, T. Behrens<sup>b</sup>, E. Ben-Dor<sup>c</sup>, D.J. Brown<sup>d</sup>, J.A.M. Demattê<sup>e</sup>, K.D. Shepherd<sup>f</sup>, Z. Shi<sup>g</sup>, B. Stenberg<sup>h</sup>, A. Stevens<sup>i</sup>, V. Adamchuk<sup>j</sup>, H. Aichi<sup>k</sup>, B.G. Barthès<sup>l</sup>, H.M. Bartholomeus<sup>m</sup>, A.D. Bayer<sup>n</sup>, M. Bernoux<sup>o</sup>, K. Böttcher<sup>op</sup>, L. Brodský<sup>q</sup>, C.W. Du<sup>r</sup>, A. Chappell<sup>s</sup>, Y. Fouad<sup>t</sup>, V. Genot<sup>u</sup>, C. Gomez<sup>v</sup>, S. Grunwald<sup>v</sup>, A. Gubler<sup>w</sup>, C. Guerrero<sup>x</sup>, C.B. Hedley<sup>y</sup>, M. Knadel<sup>z</sup>, H.J.M. Morrás<sup>aa</sup>, M. Nocita<sup>ab</sup>, L. Ramirez-Lopez<sup>ac</sup>, P. Roudier<sup>v</sup>, E.M. Rufasto Campos<sup>ad</sup>, P. Sanborn<sup>ae</sup>, V.M. Sellitto<sup>af</sup>, K.A. Sudduth<sup>ag</sup>, B.G. Rawlins<sup>ah</sup>, C. Walter<sup>ai</sup>, L.A. Winowiecki<sup>f</sup>, S.Y. Hong<sup>aj</sup>, W. Ji<sup>ak</sup>

<https://doi.org/10.1016/j.earscirev.2016.01.012>

## OSSL: Open Soil Spectral Library <https://soilspectroscopy.org/>

OSSL Open Soil Spectral Library

Soil.Spectroscopy

# of geolocated samples: 88867  
 # of MIR spectra: 45475  
 # of VNIR spectra: 45675  
 # of NIR Neospectra: 1700

Soil Security 7 (2022) 100061

Contents lists available at ScienceDirect

Soil Security

journal homepage: [www.sciencedirect.com/journal/soil-security](http://www.sciencedirect.com/journal/soil-security)

A global soil spectral calibration library and estimation service

Keith D. Shepherd<sup>a,\*</sup>, Rich Ferguson<sup>b</sup>, David Hoover<sup>b</sup>, Fenny van Egmond<sup>c</sup>, Jonathan Sanderman<sup>d</sup>, Yufeng Ge<sup>e</sup>

<sup>a</sup> Innovative Solutions for Decision Agriculture (ISDA), Rothamsted Campus, West Common, Harpenden, Hemel Hempstead AL5 2JQ, United Kingdom  
<sup>b</sup> United States Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS), Soil and Plant Science Division, National Soil Survey Center, which contains the Kellogg Soil Survey Laboratory (KSSL), Lincoln, NE, U.S.  
<sup>c</sup> ISRIC World Soil Information, Droevendaalsesteeg 3, 6708 PB, Wageningen, the Netherlands  
<sup>d</sup> Woodwell Climate Research Center, 149 Woods Hole Road, Falmouth MA 02540, USA  
<sup>e</sup> Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA



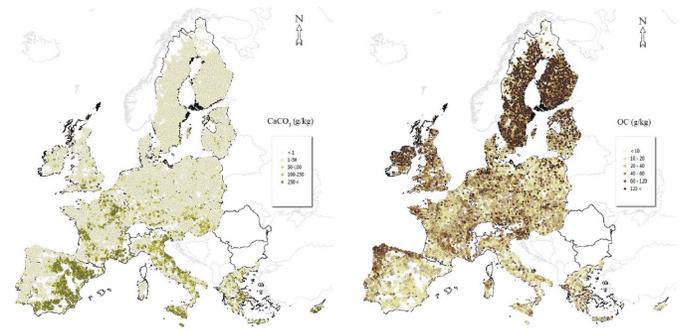
<https://doi.org/10.1016/j.soisec.2022.100061>

## LUCAS: Land Use/Cover Area frame Statistical Survey

\* Espectros vis-NIR (400-2500 nm)

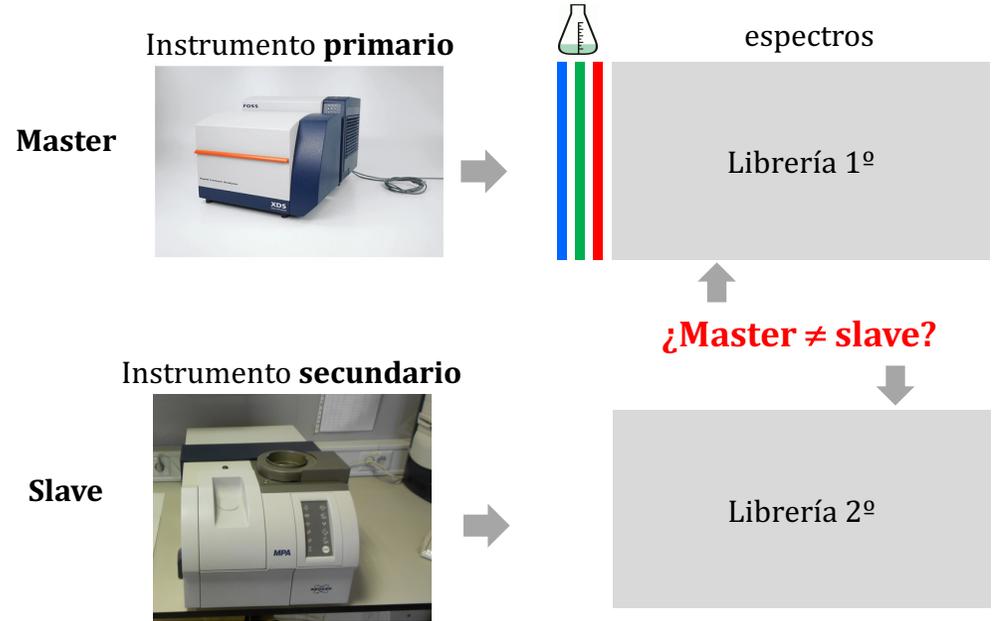
\* Varias propiedades edáficas:

- pH (H<sub>2</sub>O)
- pH (CaCl<sub>2</sub>)
- SOC
- N
- Carbonatos
- Textura...



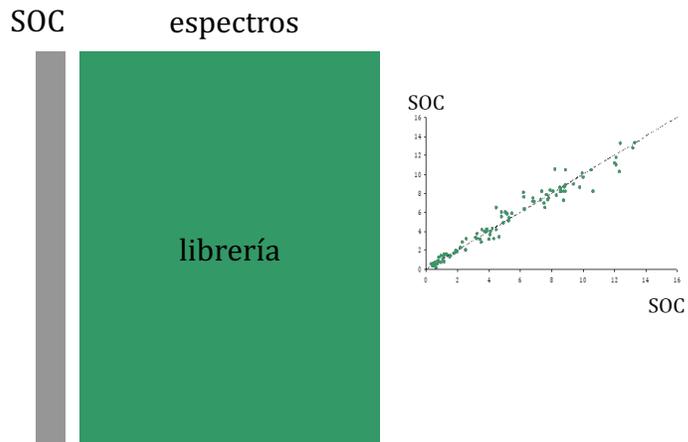
~20.000 muestras

Accesible <http://eusoils.jrc.ec.europa.eu/projects/Lucas/>

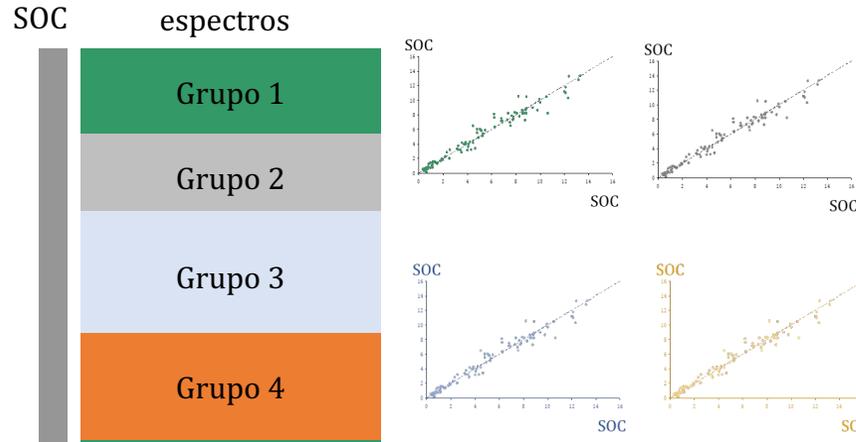


\* Técnicas de **transferencia** de calibraciones (clonar equipos)

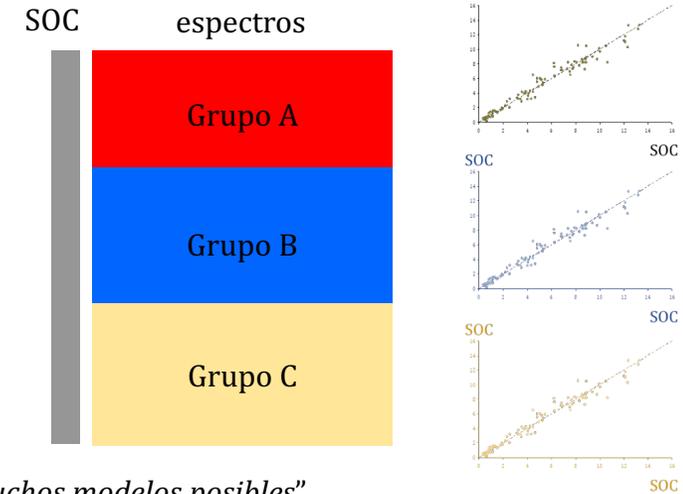
- 1.a) Escanear sub-sets en ambos instrumentos
- 1.b) Escanear "referencias"
- 2) Analizar diferencias entre instrumentos
- 3) Aplicar "correcciones" en espectros del instrumento secundario



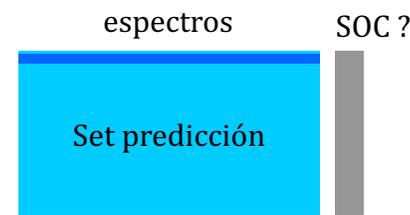
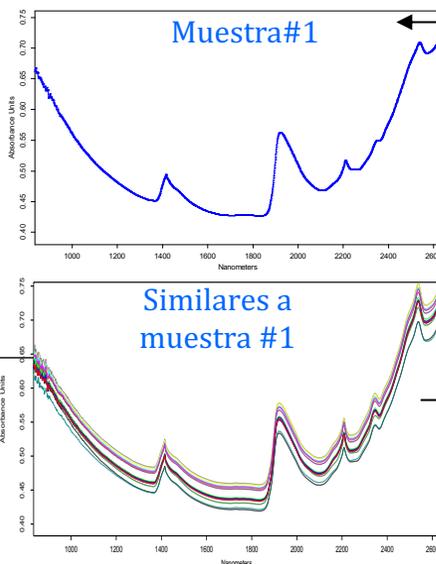
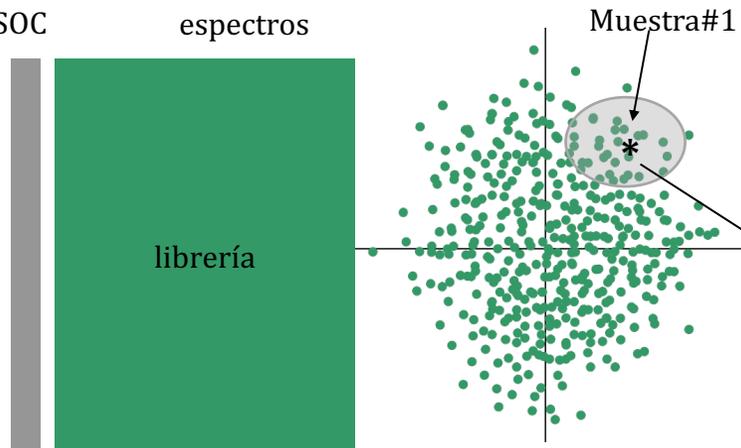
**Global:** "Una librería, un modelo"  
Idea clásica, pero "obsoleta" para suelos



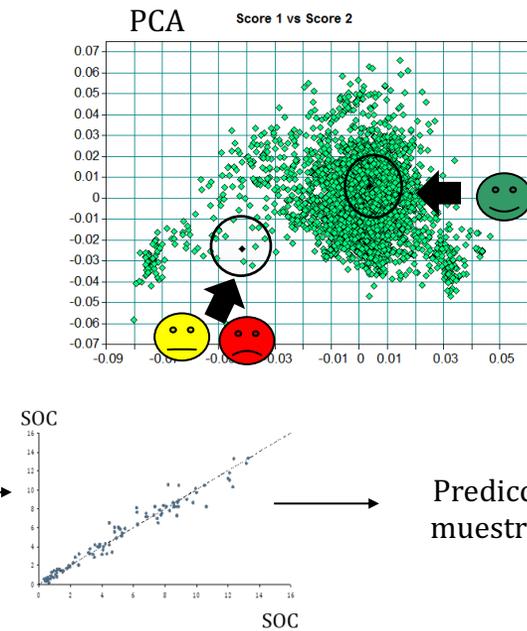
**No global:** "De una librería podemos sacar muchos modelos posibles"  
Sub-sets o grupos: geografía, litología, textura, espectral... 1º clasificas + 2º predices

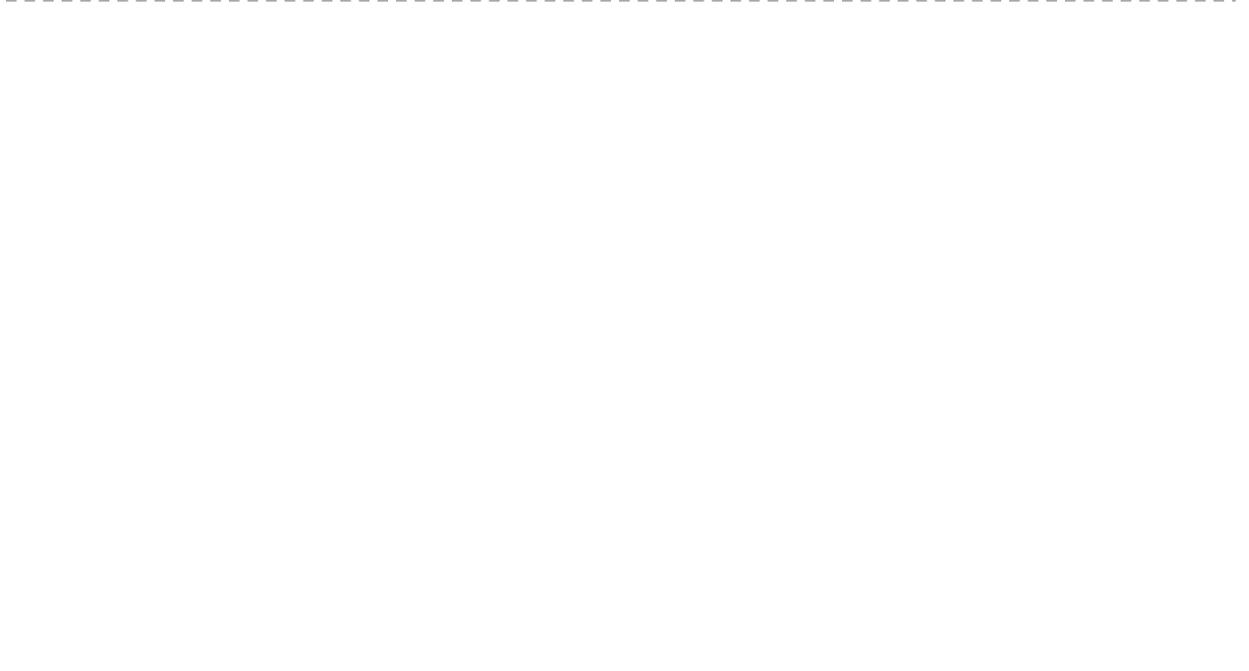
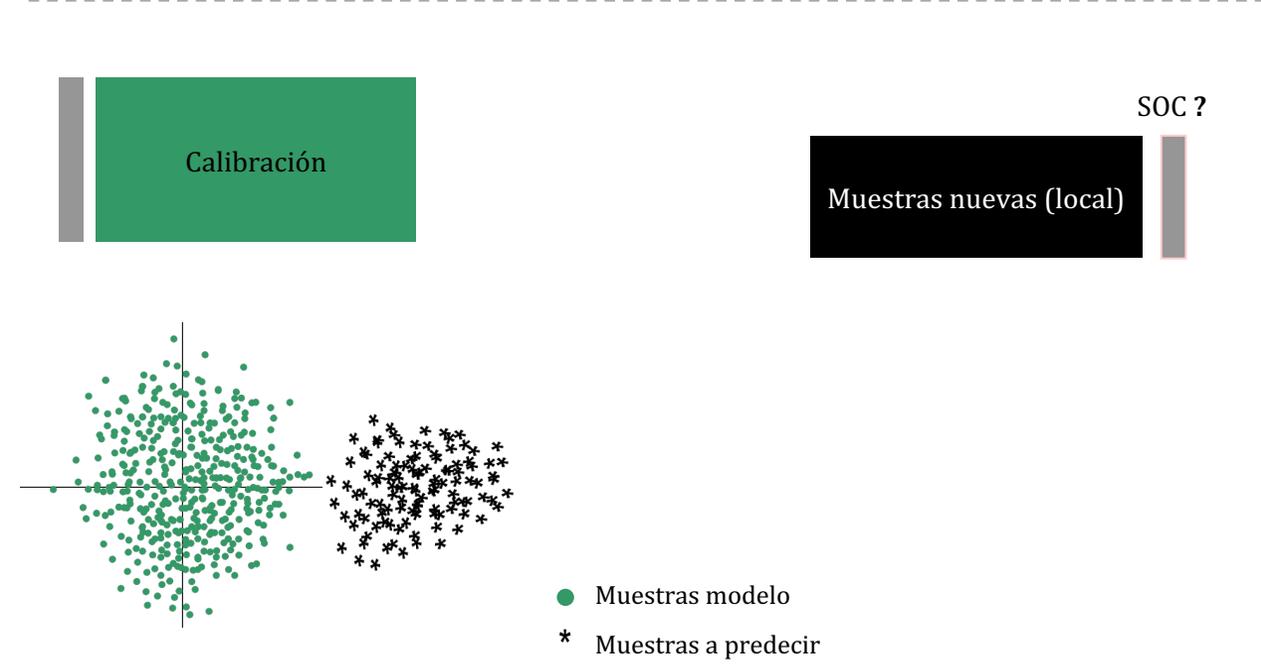
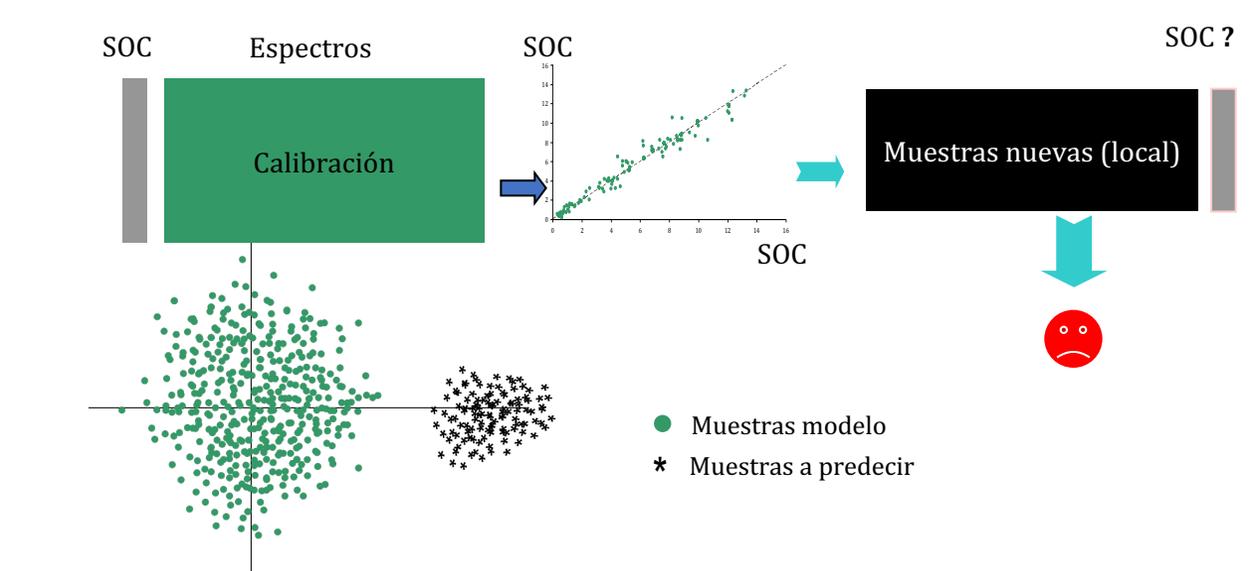
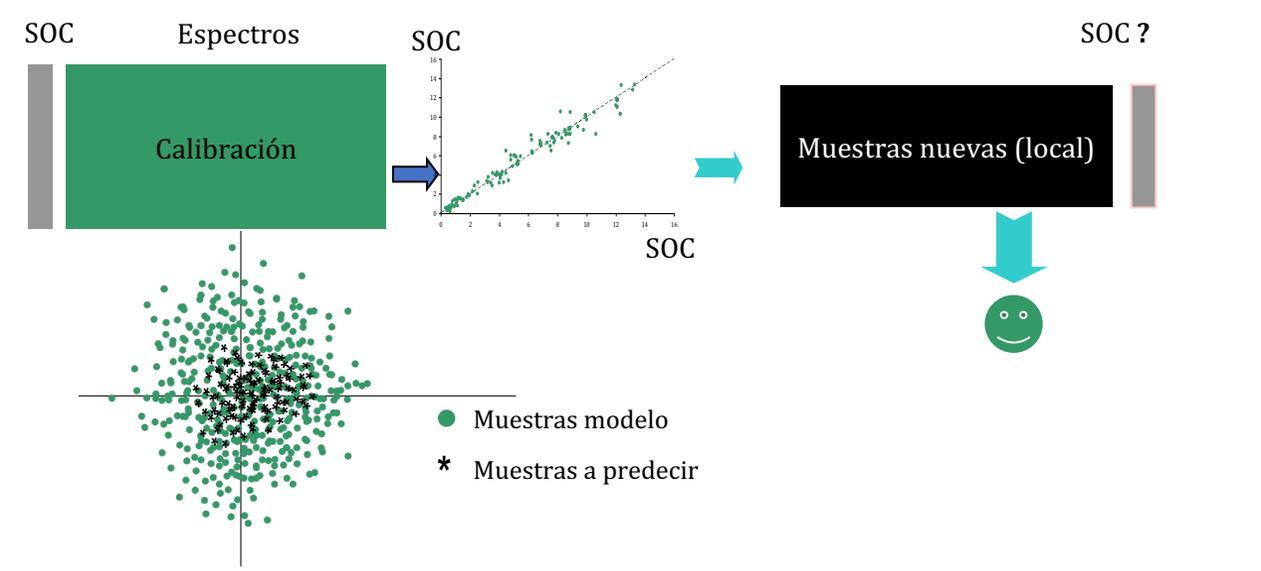


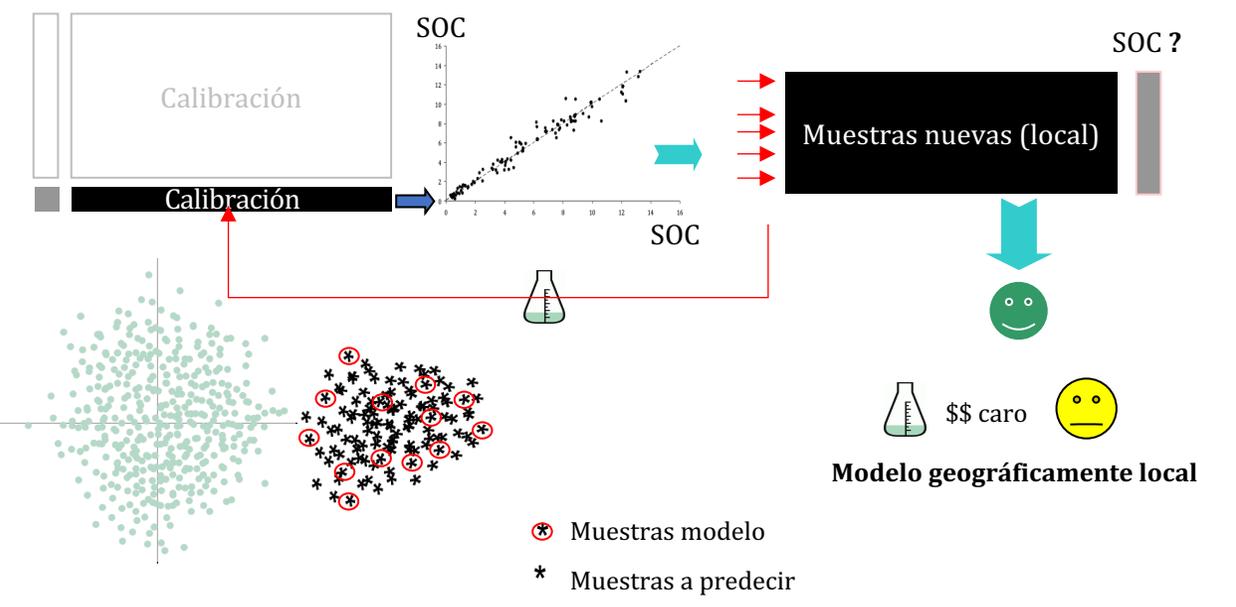
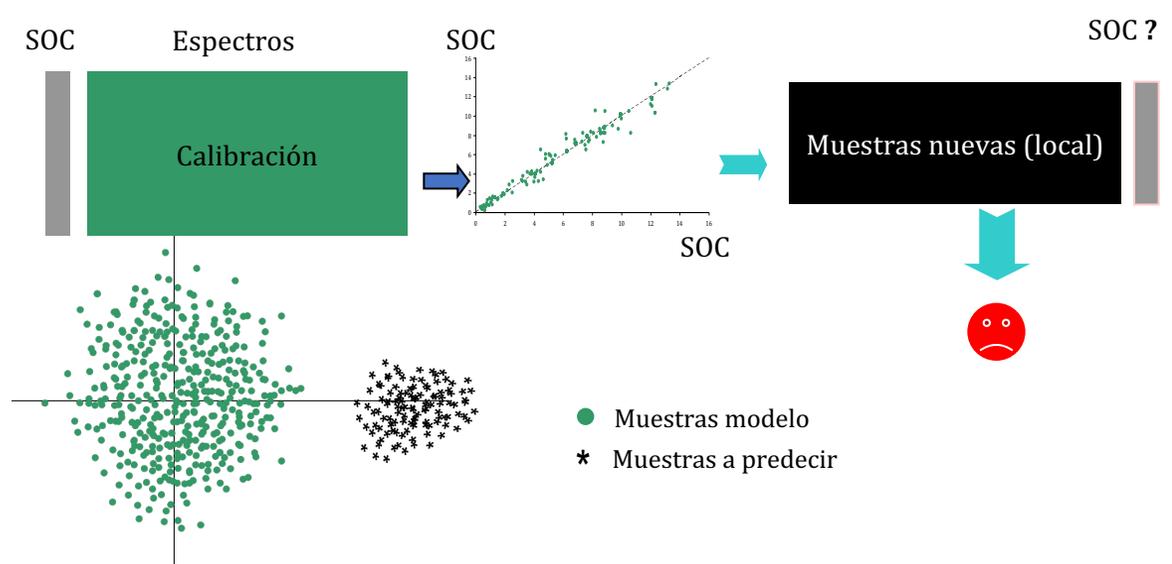
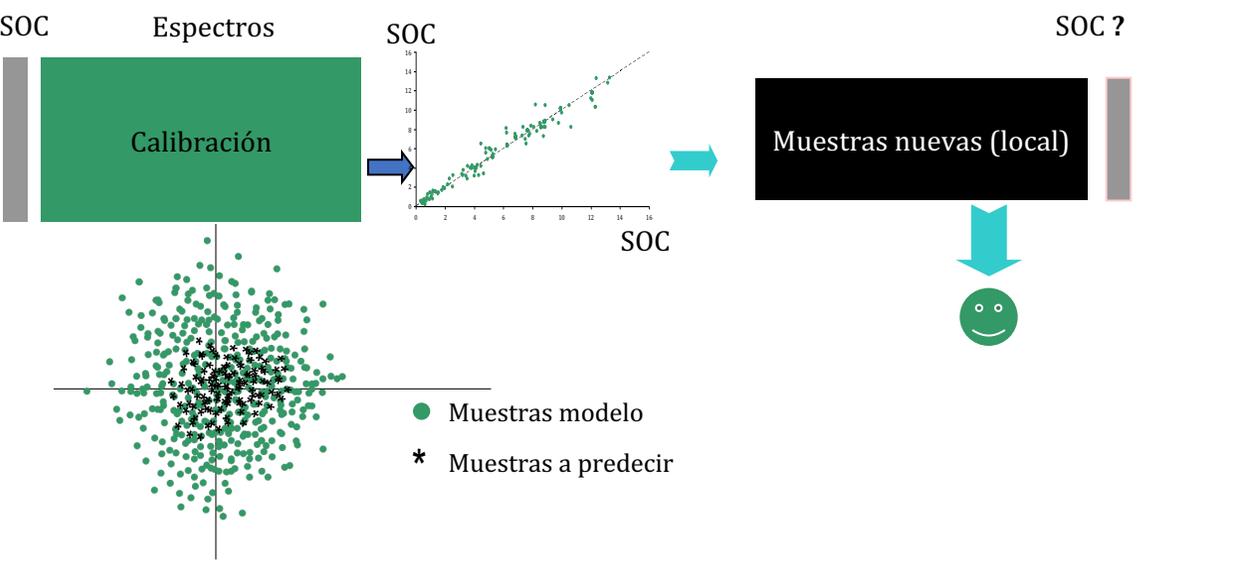
**Local:** "Modelos a medida"  
Un modelo distinto para cada muestra a predecir

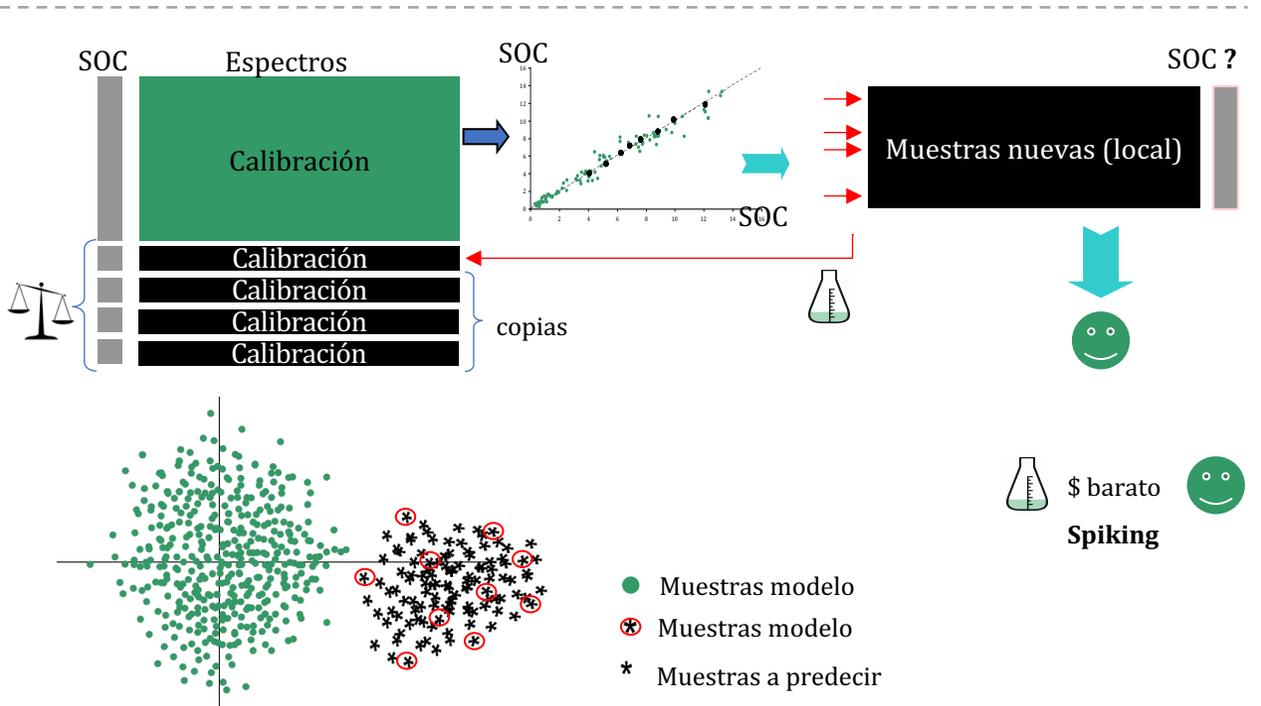
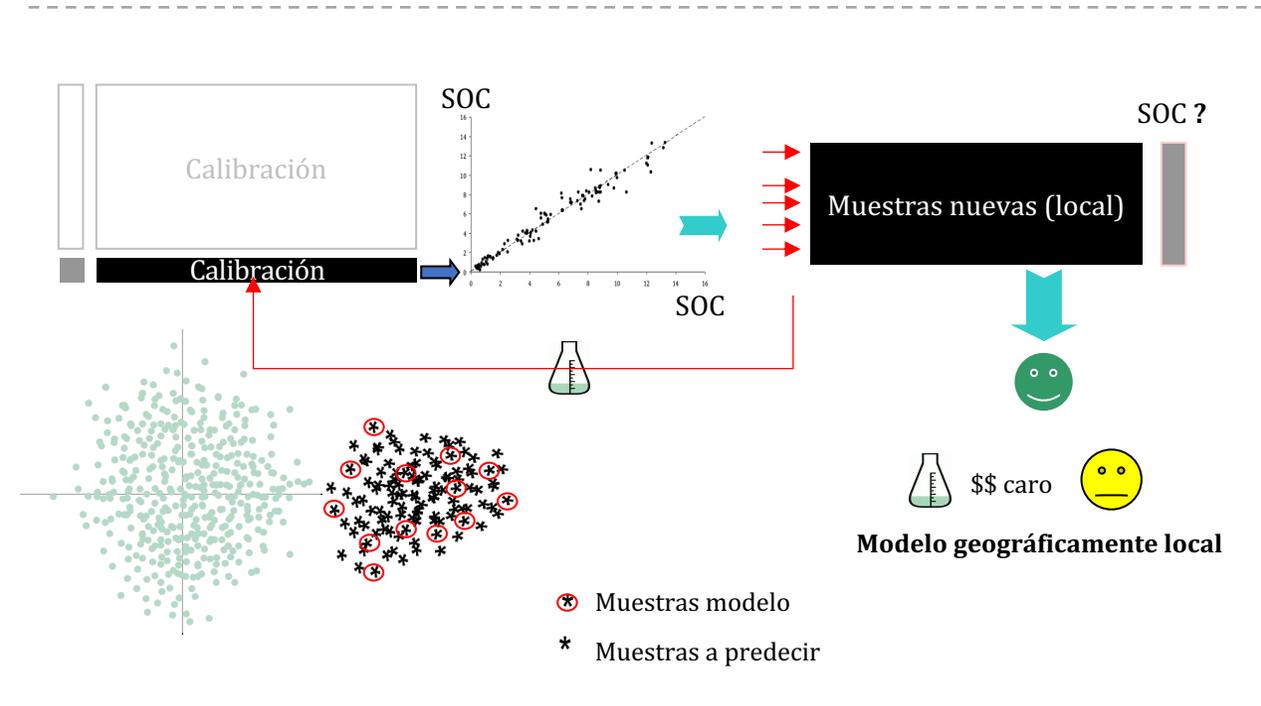
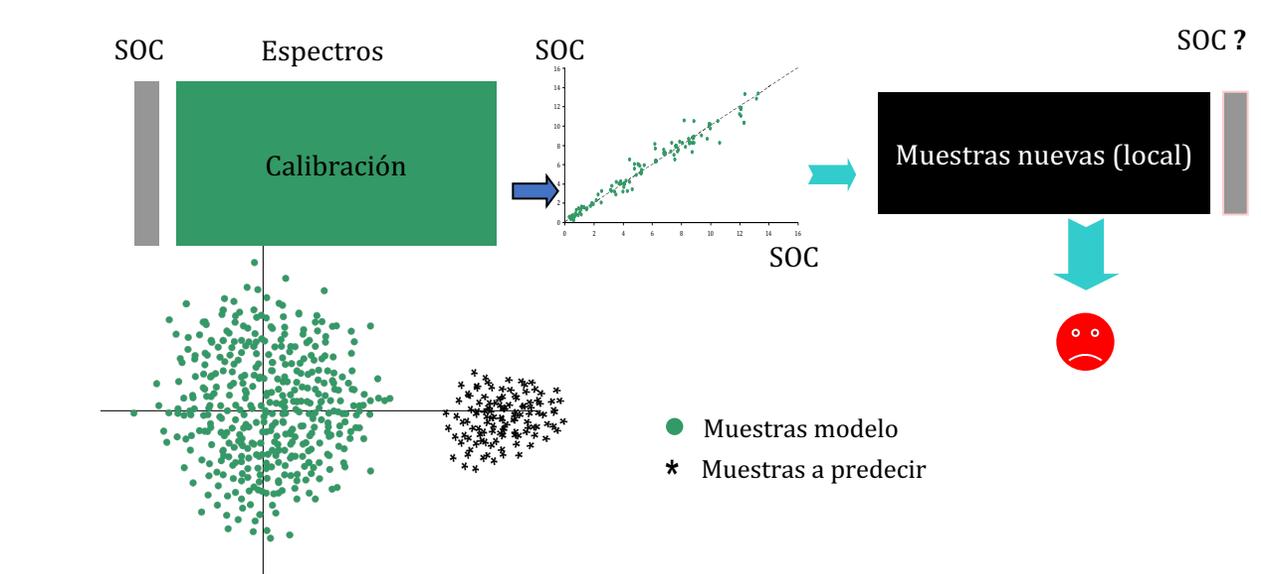
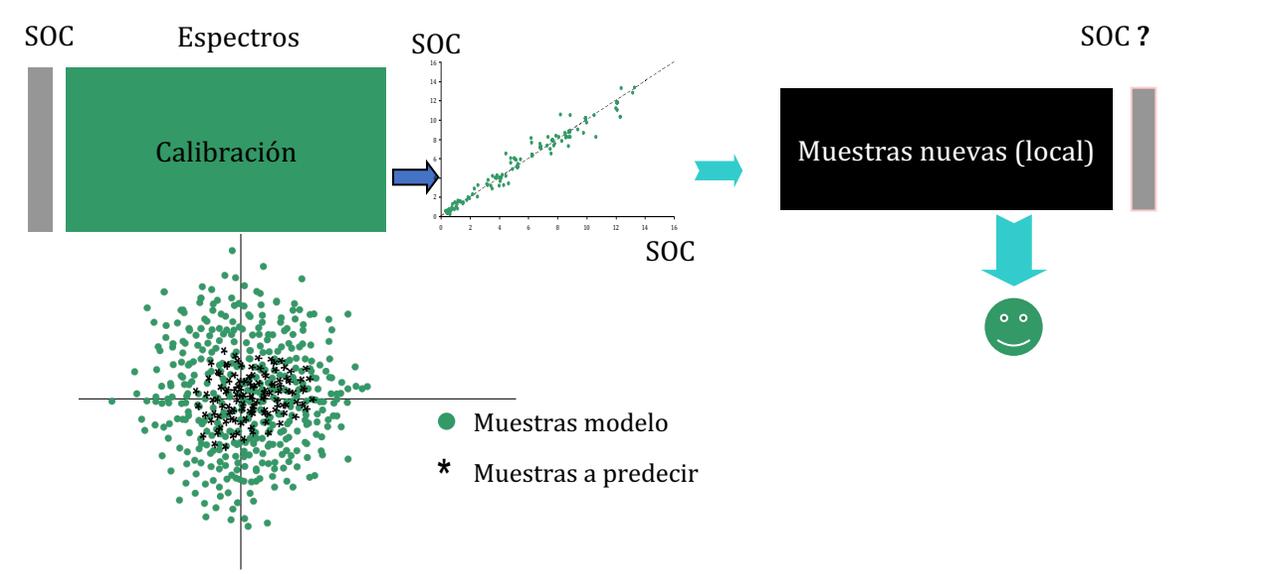


Similares a muestra #1  
Muestras para modelo de la muestra#1

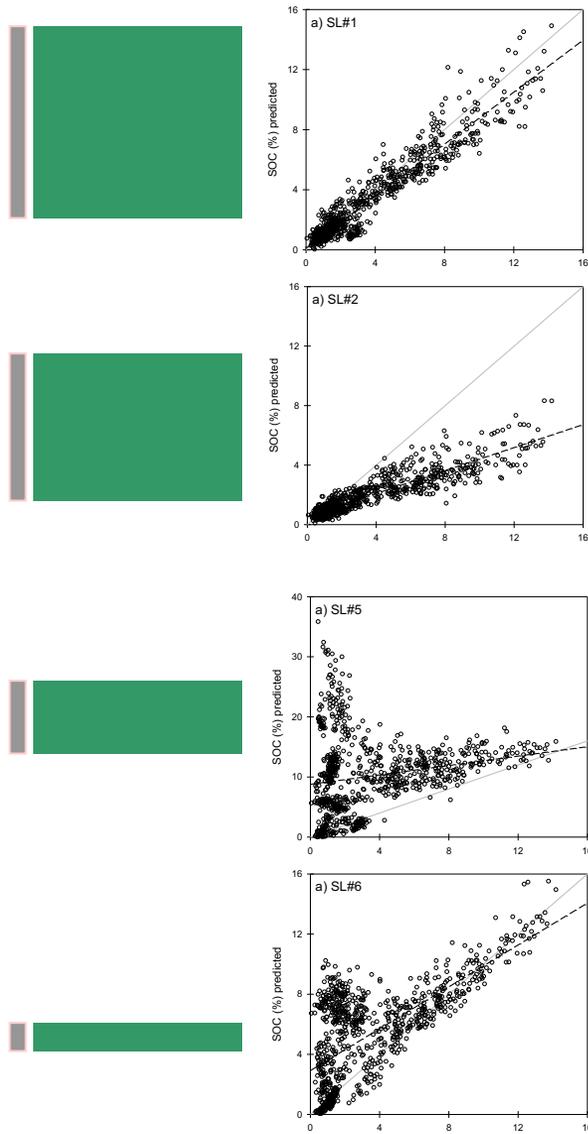




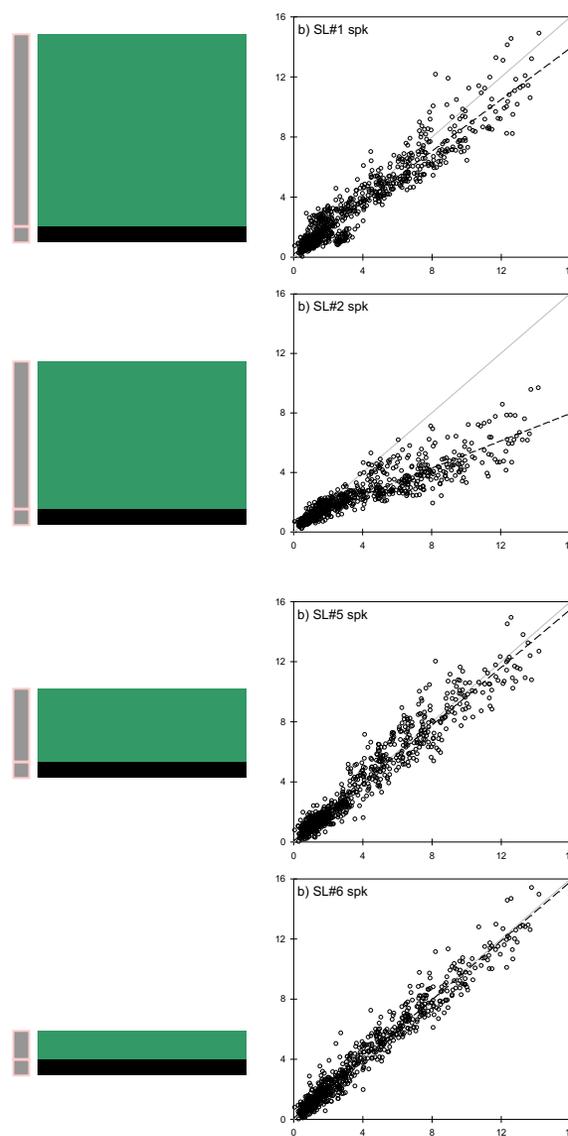




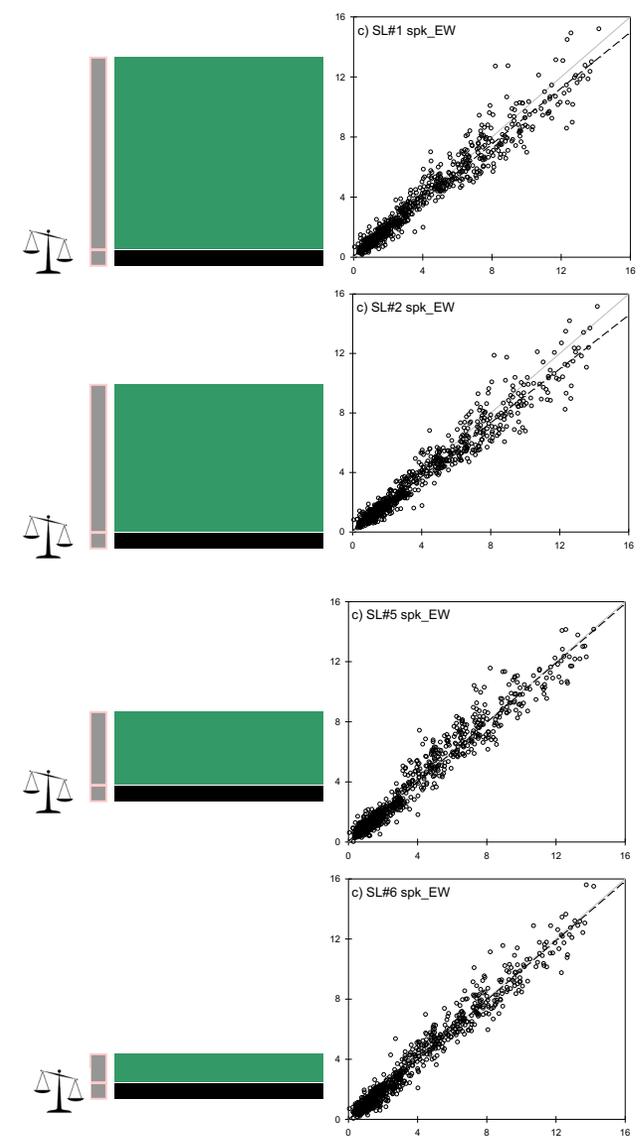
## Sólo librería



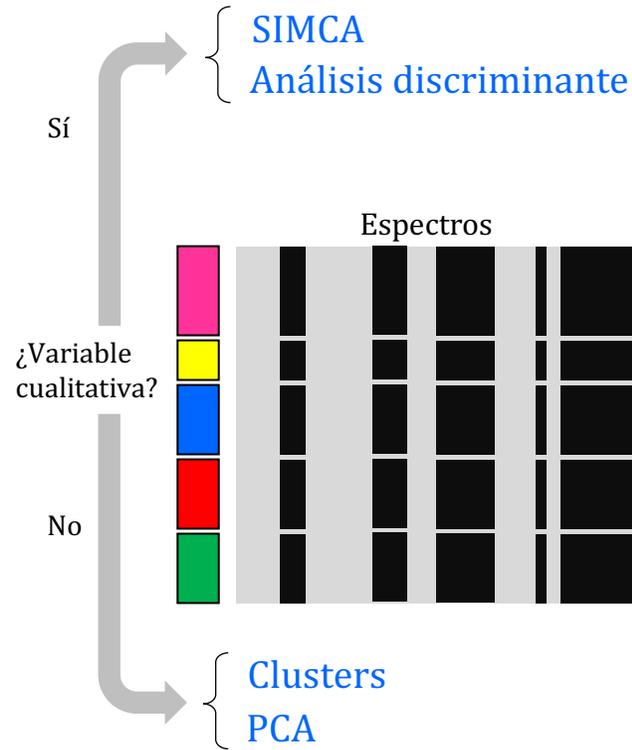
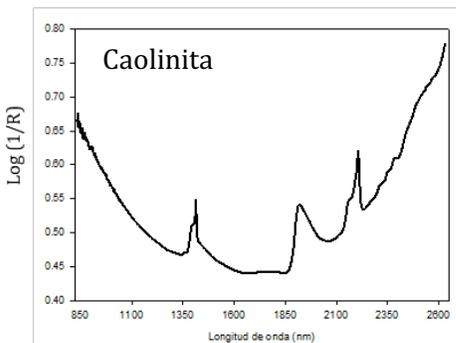
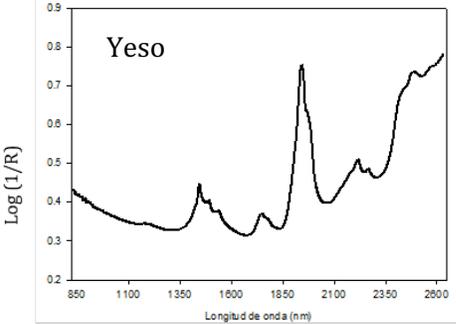
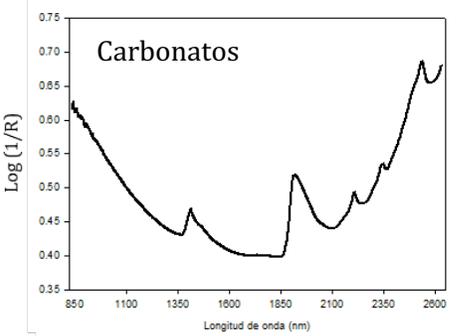
## Librería y locales (*spiking*)



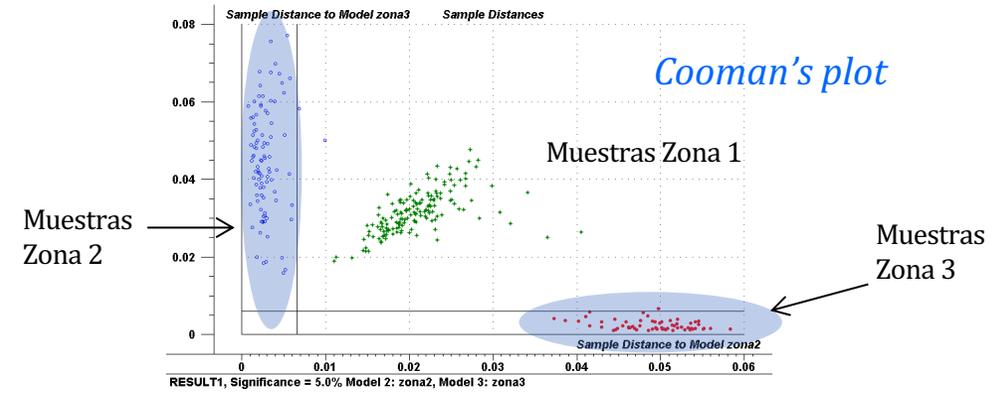
## Librería y locales (*spiking* + peso)



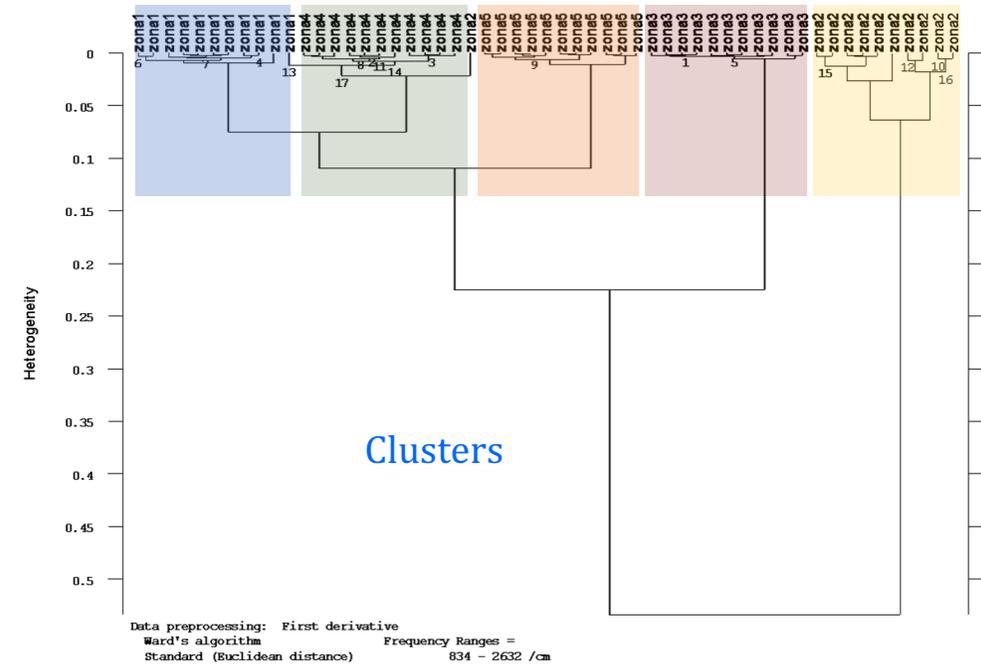
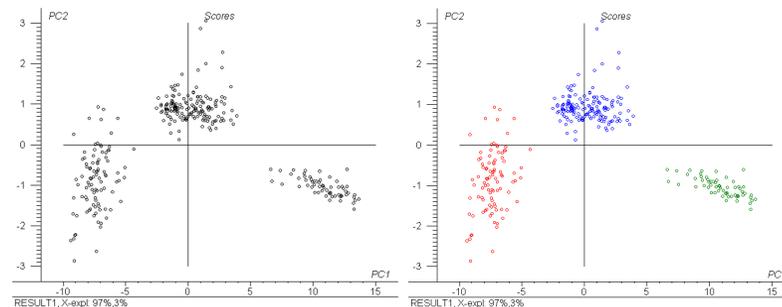
## Análisis visual minerales



## SIMCA (soft independent modeling class analogy)



## PCA (principal component analysis)



- El proceso que sirve para terminar de definir cómo es el modelo no debe interpretarse como una verdadera validación.
- Diseña todas tus pruebas para que simulen el uso real del modelo:
  - ✓ Posiblemente los resultados no serán tan “aparentemente” buenos, pero serán realistas: necesario para conocer la verdadera capacidad de esta técnica
  - ✓ Te ayudará a conocer las fortalezas y debilidades de los modelos: identificar cómo mejorarlos
- Algunos resultados reportados en bibliografía se han obtenido sin usar sets independientes (de manera no consciente): lee con cautela y “críticamente”.
- No todas las propiedades edáficas tienen respuesta espectral (no vale para analizar todo), ni se trata de reemplazar a los métodos clásicos al 100% (aproximaciones complementarias).
- Analiza los resultados poniéndolos en contexto de aplicación (no simplemente mires los valores de  $R^2$  o RMSEP). Compáralos con los obtenidos en tus pruebas de reproducibilidad (réplicas del método de referencia en días distintos)
- Existen otros algoritmos además de PLS (evita “*black-box*”)
- “*Garbage in, garbage out*”: mantén la calidad de (i) los espectros (instrumento), (ii) datos analíticos, (iii) modelos
- Reflexiona: ¿Cuál va a ser el uso del modelo? ¿escala local? ¿dispones de una gran librería? Elige la mejor estrategia, e identifica dónde invertir tus esfuerzos.