Interreg

España – Portugal

Cofinanciado por la Unión Europea

Cofinanciado pela União Europeia

TID4AGRO

Desarrollo de modelos predictivos mediante tecnología NIRS

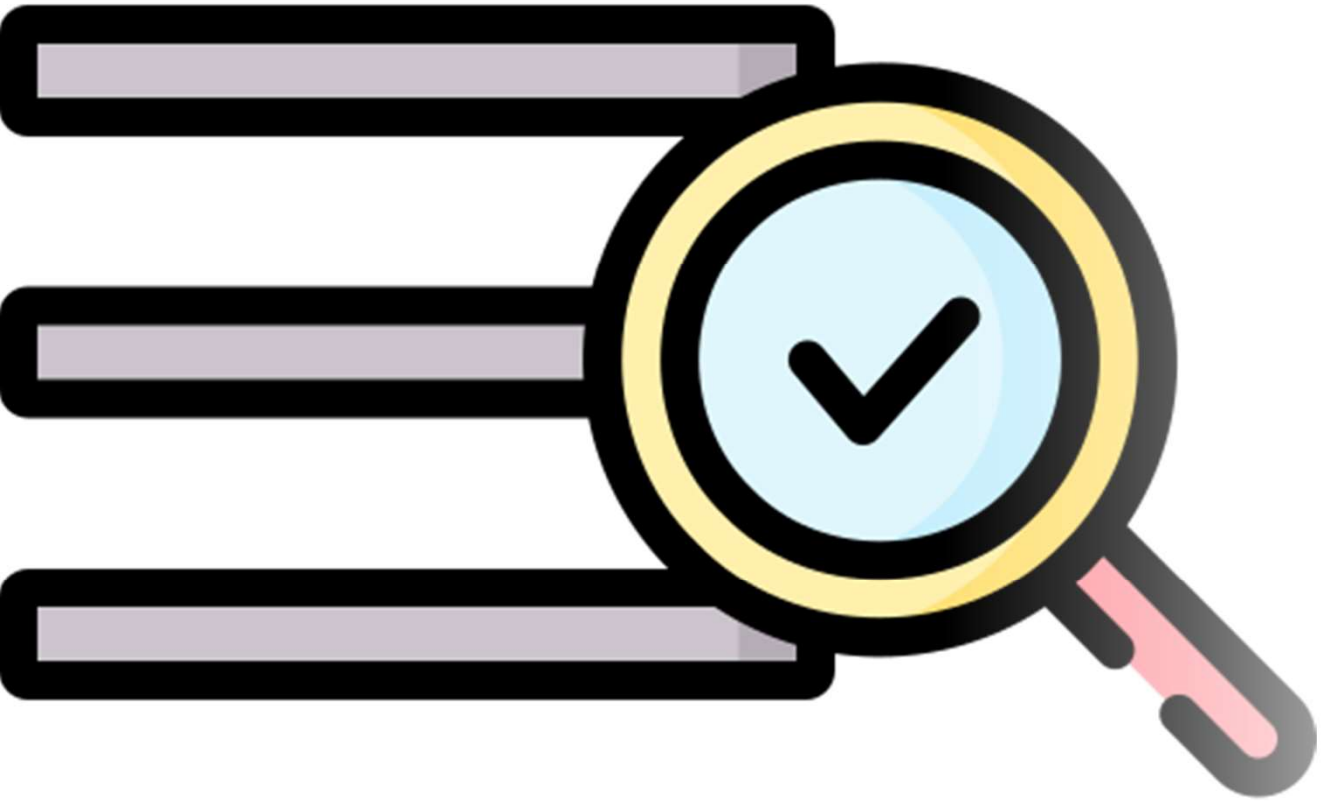
Tecnologías avanzadas, innovadoras y digitales para el sector agroalimentario de la EUROACE

Évora, 18th and 19th, novembre, 2024

Researcher of the Meat Quality Area of CICYTEX
PhD in Food Science
alberto.ortiz@juntaex.es

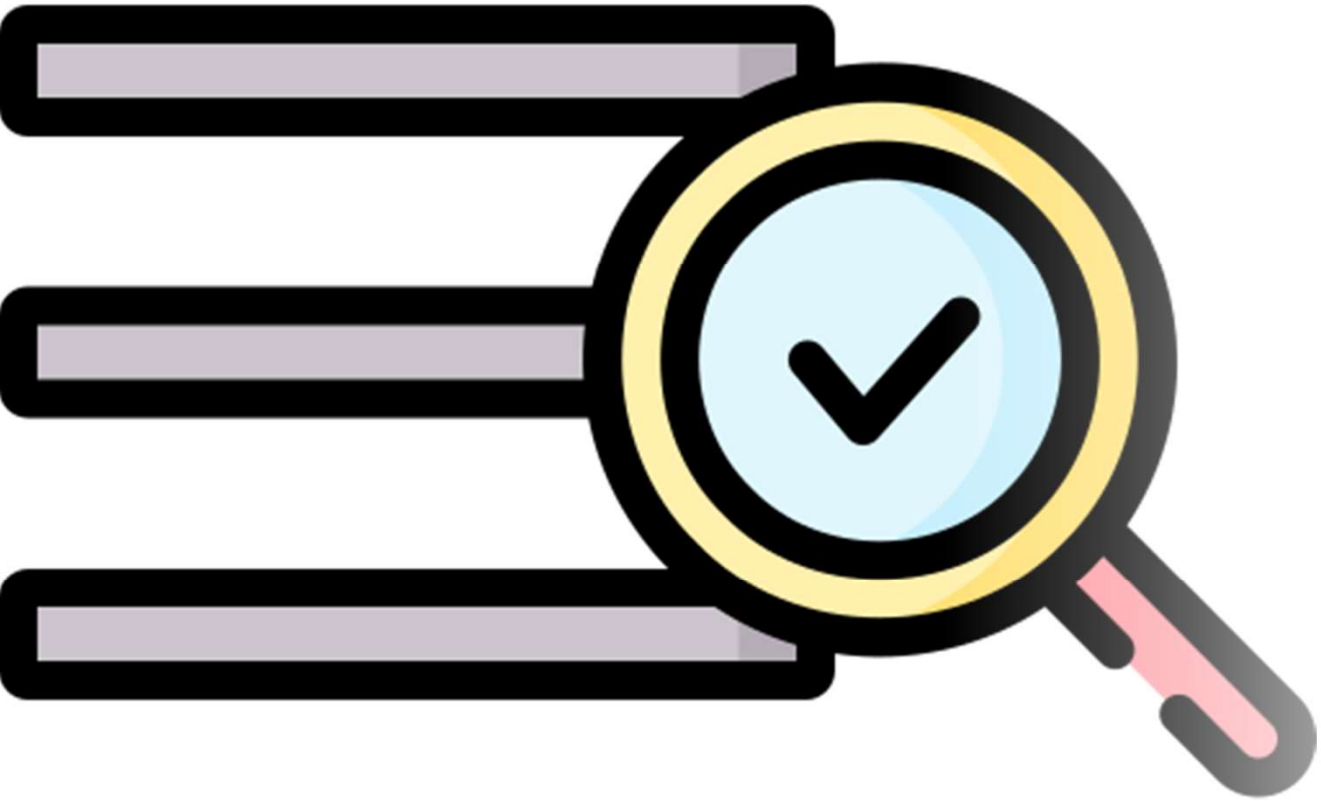
Degree in food science and technology
Meat Quality Area of CICYTEX
lucia.leon@juntaex.es

Centro de Investigaciones Científicas y Tecnológicas de Extremadura, CICYTEX.
A5, km 472. 06187 Guadajira (Badajoz)



Index

- What is NIRS technology?
- The Unscrambler
  - The construction of the matrix
  - Quantitative models
  - Qualitative models
  - Pre-treatments



Index

- **What is NIRS technology?**
- The Unscrambler
  - The construction of the matrix
  - Quantitative models
  - Qualitative models
  - Pre-treatments

What is NIRS technology?

What is NIRS technology?

Near

InfraRed

Spectroscopy

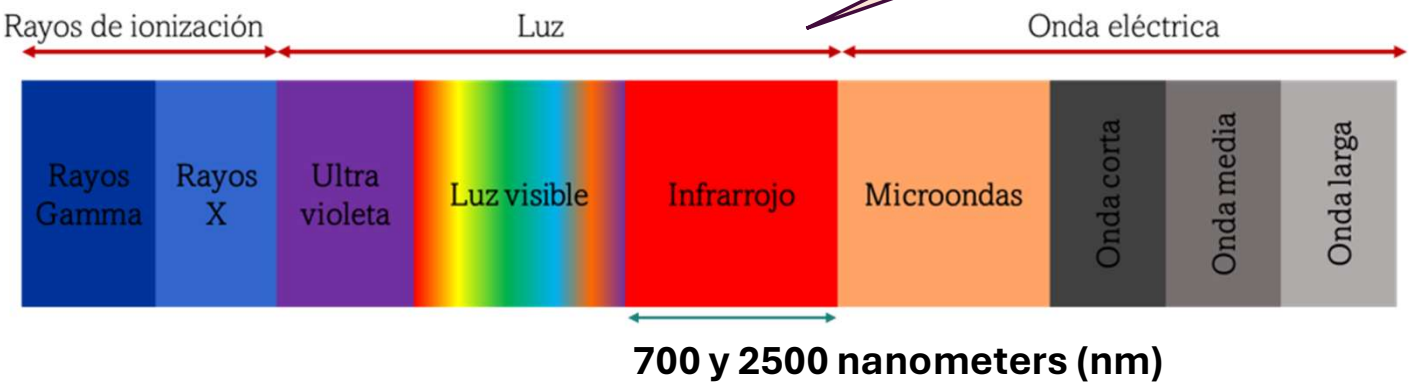
What is NIRS technology?

Near

InfraRed

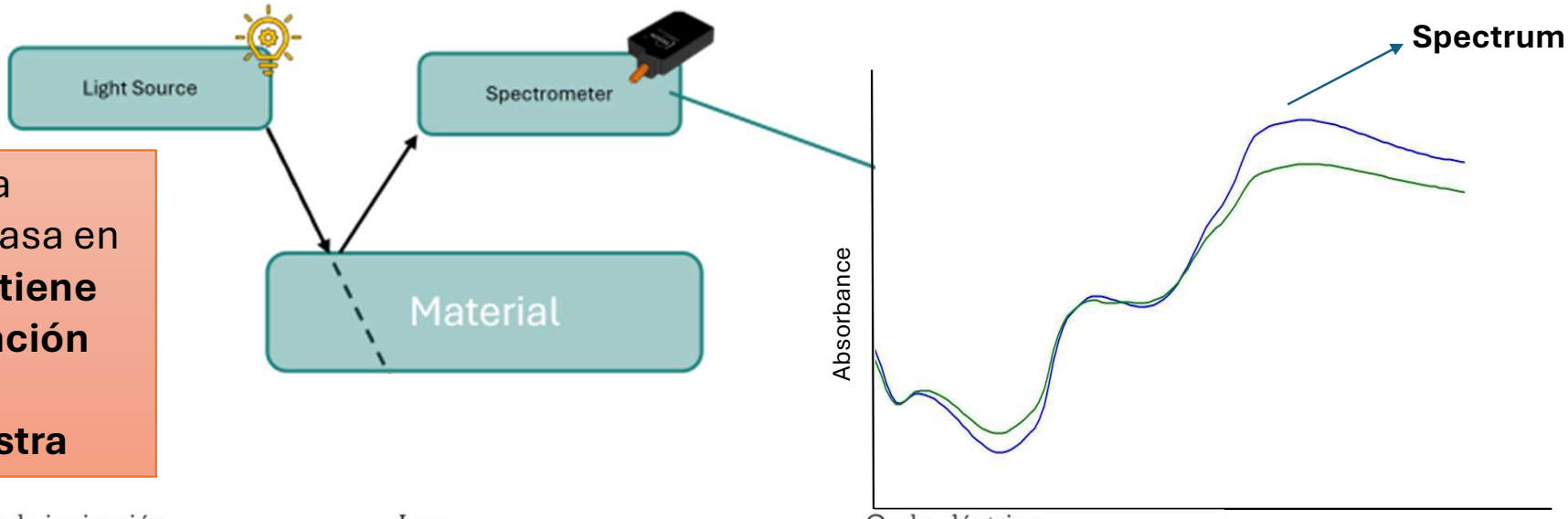
Spectroscopy

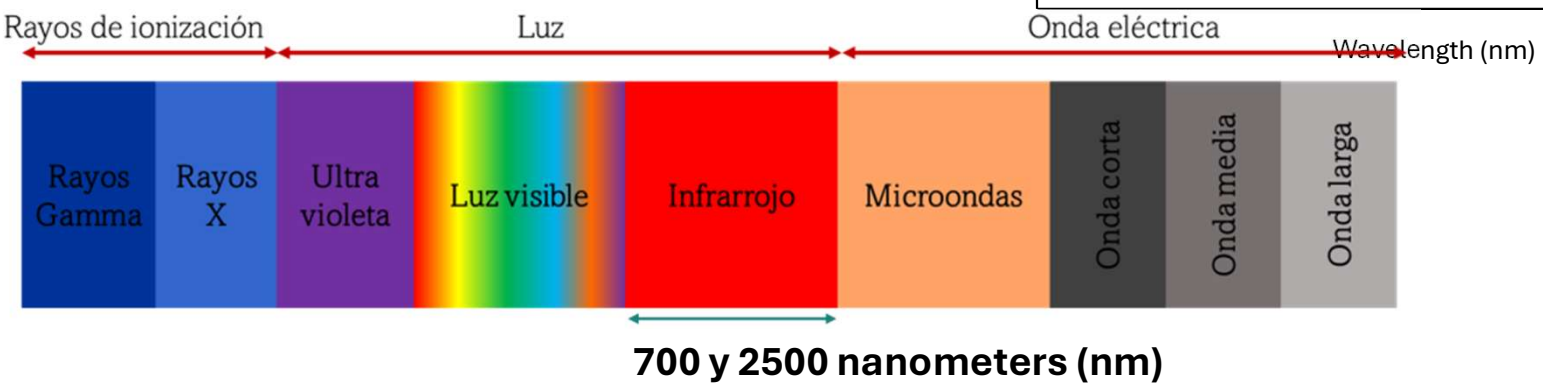
Near Infrared Spectroscopy (NIR) is an **analytical technique** that uses the **near infrared region** of the electromagnetic spectrum.



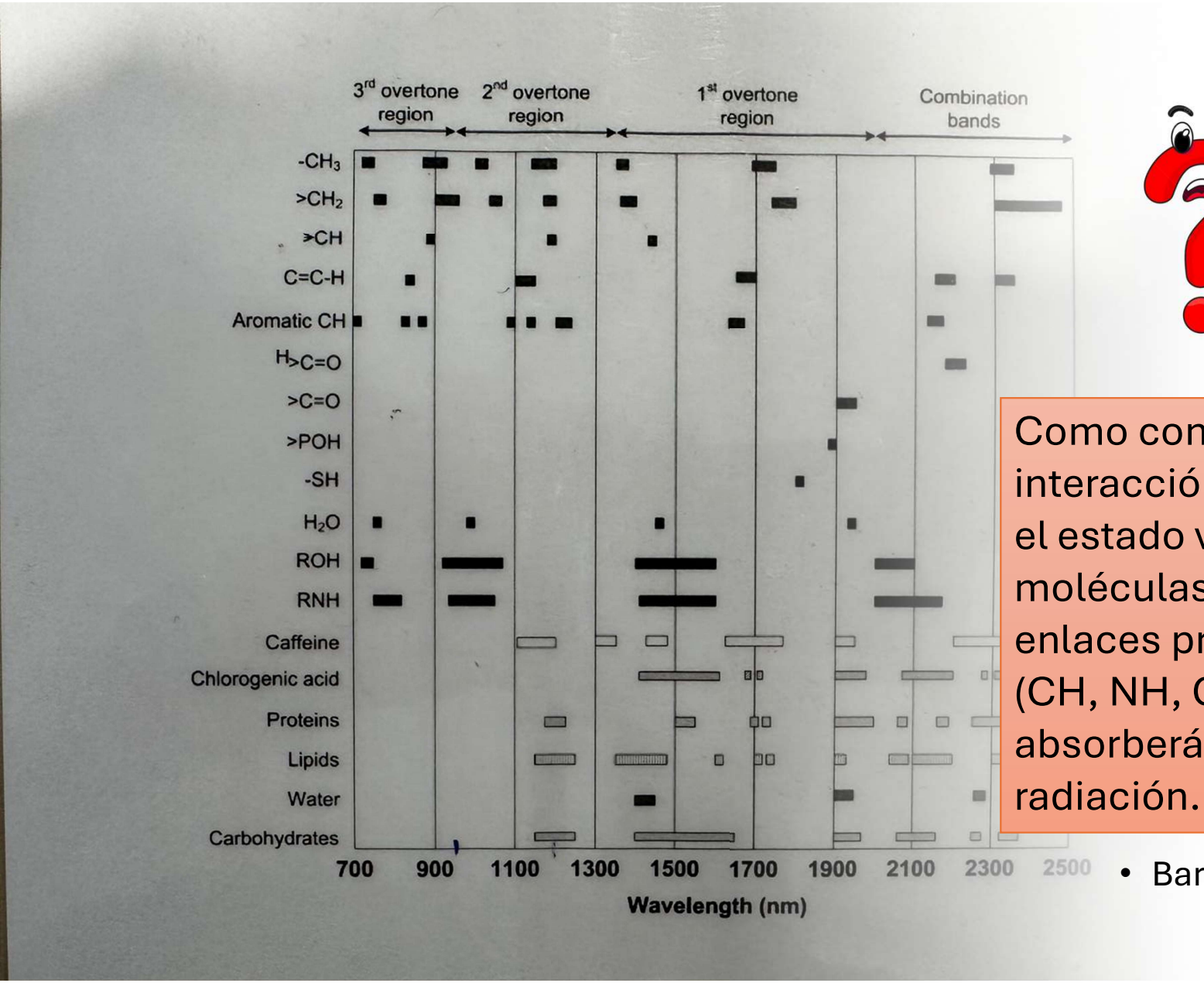
What is NIRS technology?

El fundamento de la tecnología NIR se basa en la **interacción que tiene lugar entre la radiación electromagnética infrarroja y la muestra**





700 y 2500 nanometers (nm)



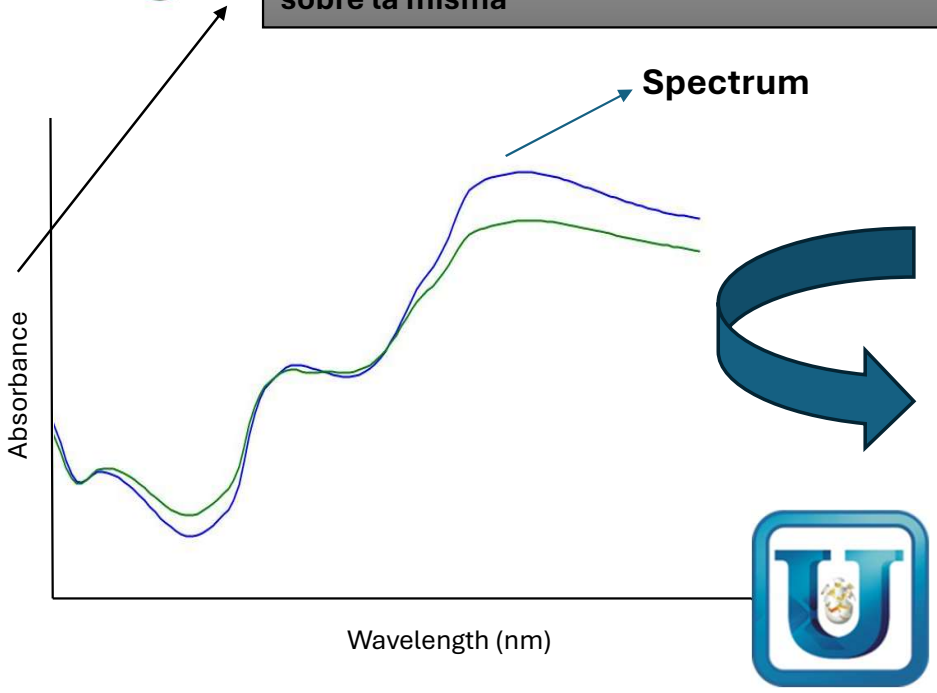
What is NIRS technology?

Como consecuencia de esta interacción se producen cambios en el estado vibracional de las moléculas. Así, en función de los enlaces presentes en las moléculas (CH, NH, OH, CO) de la muestra, se absorberá una determinada radiación.

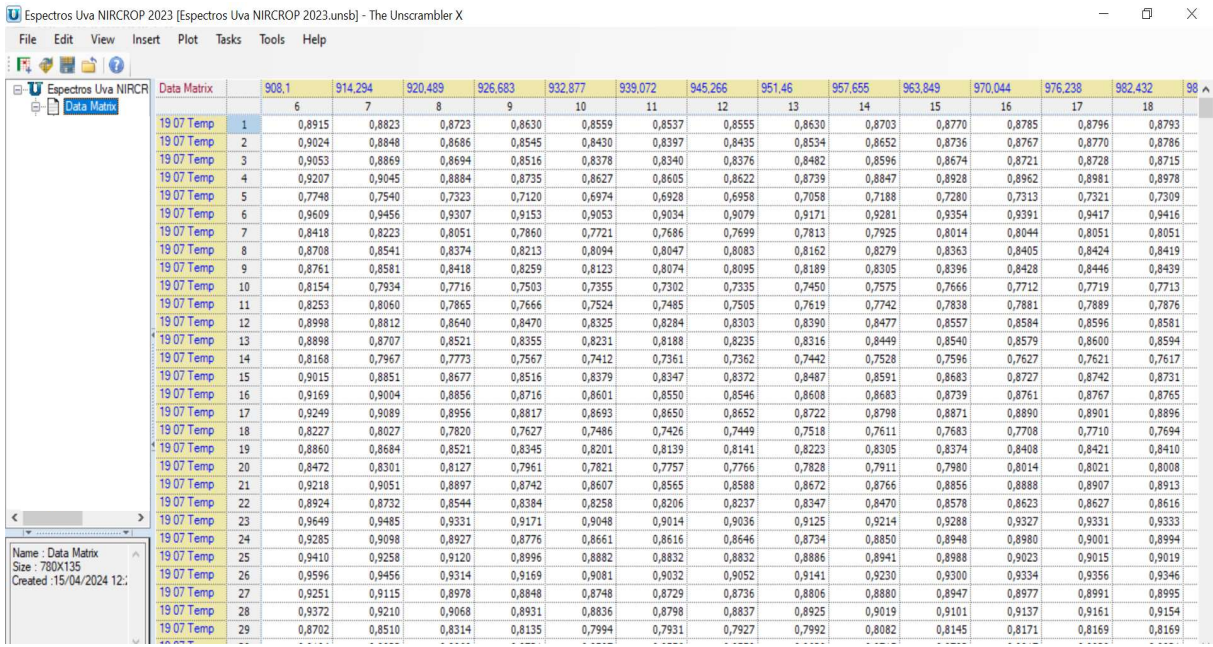
- Barbin et al., 2014

What is NIRS technology?

Absorbancia = Log (1/R), donde R son los valores de Reflectancia o **cociente entre la radiación reflejada por la muestra y la incidente sobre la misma**



Absorbance values



Data Matrix		908,1	914,294	920,489	926,683	932,877	939,072	945,266	951,46	957,655	963,849	970,044	976,238	982,432	98
		6	7	8	9	10	11	12	13	14	15	16	17	18	
19 07 Temp	1	0,8915	0,8823	0,8723	0,8630	0,8559	0,8537	0,8555	0,8630	0,8703	0,8770	0,8785	0,8796	0,8793	
19 07 Temp	2	0,9024	0,8848	0,8686	0,8545	0,8430	0,8397	0,8435	0,8534	0,8652	0,8736	0,8767	0,8770	0,8786	
19 07 Temp	3	0,9053	0,8869	0,8694	0,8516	0,8378	0,8340	0,8376	0,8482	0,8596	0,8674	0,8721	0,8728	0,8715	
19 07 Temp	4	0,9207	0,9045	0,8884	0,8735	0,8627	0,8605	0,8622	0,8739	0,8847	0,8928	0,8962	0,8981	0,8978	
19 07 Temp	5	0,7748	0,7540	0,7323	0,7120	0,6974	0,6928	0,6958	0,7058	0,7188	0,7280	0,7313	0,7321	0,7309	
19 07 Temp	6	0,9609	0,9456	0,9307	0,9153	0,9053	0,9034	0,9079	0,9171	0,9281	0,9354	0,9391	0,9417	0,9416	
19 07 Temp	7	0,8418	0,8223	0,8051	0,7860	0,7721	0,7686	0,7699	0,7813	0,7925	0,8014	0,8044	0,8051	0,8051	
19 07 Temp	8	0,8708	0,8541	0,8374	0,8213	0,8094	0,8047	0,8083	0,8162	0,8279	0,8363	0,8405	0,8424	0,8419	
19 07 Temp	9	0,8761	0,8581	0,8418	0,8259	0,8123	0,8074	0,8095	0,8189	0,8305	0,8396	0,8428	0,8446	0,8439	
19 07 Temp	10	0,8154	0,7934	0,7716	0,7503	0,7355	0,7302	0,7335	0,7450	0,7575	0,7666	0,7712	0,7719	0,7713	
19 07 Temp	11	0,8253	0,8060	0,7865	0,7666	0,7524	0,7485	0,7505	0,7619	0,7742	0,7838	0,7881	0,7889	0,7876	
19 07 Temp	12	0,8998	0,8812	0,8640	0,8470	0,8325	0,8284	0,8303	0,8390	0,8477	0,8557	0,8584	0,8596	0,8581	
19 07 Temp	13	0,8898	0,8707	0,8521	0,8355	0,8231	0,8188	0,8235	0,8316	0,8449	0,8540	0,8579	0,8600	0,8594	
19 07 Temp	14	0,8168	0,7967	0,7773	0,7567	0,7412	0,7361	0,7362	0,7442	0,7528	0,7596	0,7627	0,7621	0,7617	
19 07 Temp	15	0,9015	0,8851	0,8677	0,8516	0,8379	0,8347	0,8372	0,8487	0,8591	0,8683	0,8727	0,8742	0,8731	
19 07 Temp	16	0,9169	0,9004	0,8856	0,8716	0,8601	0,8550	0,8546	0,8608	0,8683	0,8739	0,8761	0,8767	0,8765	
19 07 Temp	17	0,9249	0,9089	0,8956	0,8817	0,8693	0,8650	0,8652	0,8722	0,8798	0,8871	0,8890	0,8901	0,8896	
19 07 Temp	18	0,8227	0,8027	0,7820	0,7627	0,7486	0,7426	0,7449	0,7518	0,7611	0,7683	0,7708	0,7710	0,7694	
19 07 Temp	19	0,8860	0,8684	0,8521	0,8345	0,8201	0,8139	0,8141	0,8223	0,8305	0,8374	0,8408	0,8421	0,8410	
19 07 Temp	20	0,8472	0,8301	0,8127	0,7961	0,7821	0,7757	0,7766	0,7828	0,7911	0,7980	0,8014	0,8021	0,8008	
19 07 Temp	21	0,9218	0,9051	0,8897	0,8742	0,8607	0,8565	0,8588	0,8672	0,8766	0,8856	0,8888	0,8907	0,8913	
19 07 Temp	22	0,8924	0,8732	0,8544	0,8384	0,8258	0,8206	0,8237	0,8347	0,8470	0,8578	0,8623	0,8627	0,8616	
19 07 Temp	23	0,9649	0,9485	0,9331	0,9171	0,9048	0,9014	0,9036	0,9125	0,9214	0,9288	0,9327	0,9331	0,9333	
19 07 Temp	24	0,9285	0,9098	0,8927	0,8776	0,8661	0,8616	0,8646	0,8734	0,8850	0,8948	0,8980	0,9001	0,8994	
19 07 Temp	25	0,9410	0,9258	0,9120	0,8996	0,8882	0,8832	0,8832	0,8886	0,8941	0,8988	0,9023	0,9015	0,9019	
19 07 Temp	26	0,9596	0,9456	0,9314	0,9169	0,9081	0,9032	0,9052	0,9141	0,9230	0,9300	0,9334	0,9356	0,9346	
19 07 Temp	27	0,9251	0,9115	0,8978	0,8848	0,8748	0,8729	0,8736	0,8806	0,8880	0,8947	0,8977	0,8991	0,8995	
19 07 Temp	28	0,9372	0,9210	0,9068	0,8931	0,8836	0,8798	0,8837	0,8925	0,9019	0,9101	0,9137	0,9161	0,9154	
19 07 Temp	29	0,8702	0,8510	0,8314	0,8135	0,7994	0,7931	0,7927	0,7992	0,8082	0,8145	0,8171	0,8169	0,8169	

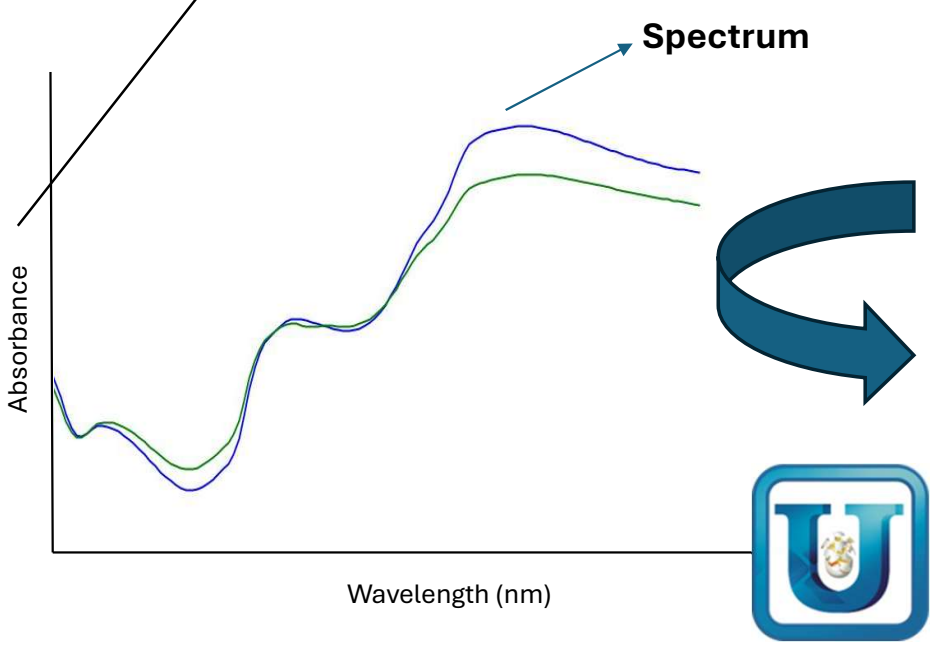
The Unscrambler

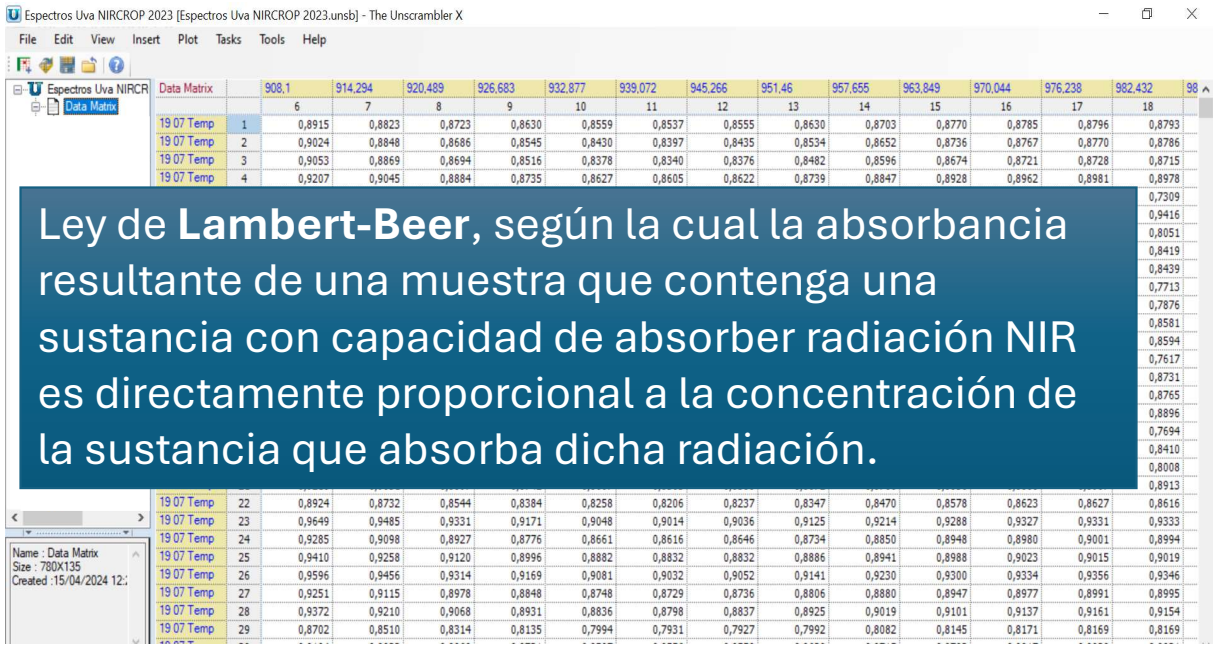
CAMO software for analytical modeling and prediction

What is NIRS technology?

Absorbancia = Log (1/R), donde R son los valores de Reflectancia o **cociente entre la radiación reflejada por la muestra y la incidente sobre la misma**

Absorbance values

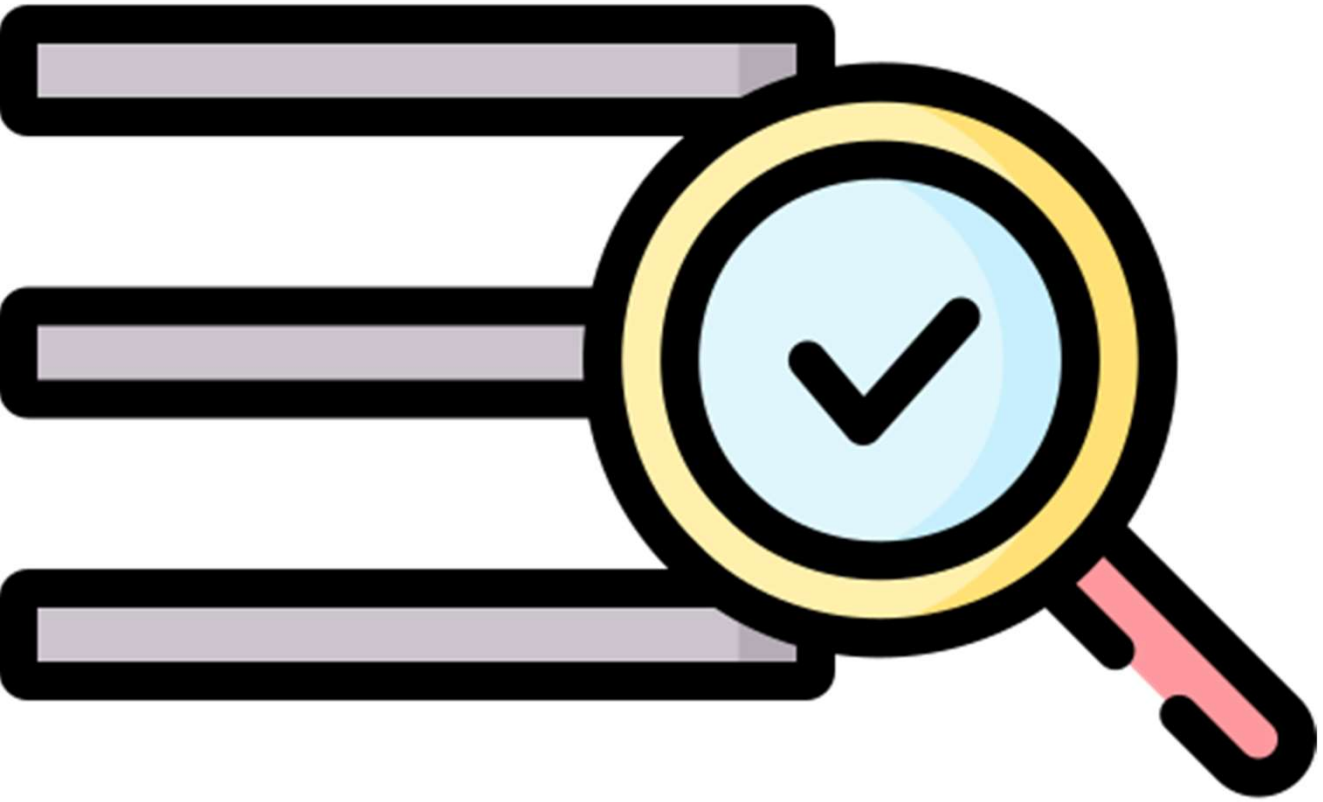




Ley de **Lambert-Beer**, según la cual la absorbancia resultante de una muestra que contenga una sustancia con capacidad de absorber radiación NIR es directamente proporcional a la concentración de la sustancia que absorba dicha radiación.

The Unscrambler

CAMO software for analytical modeling and prediction

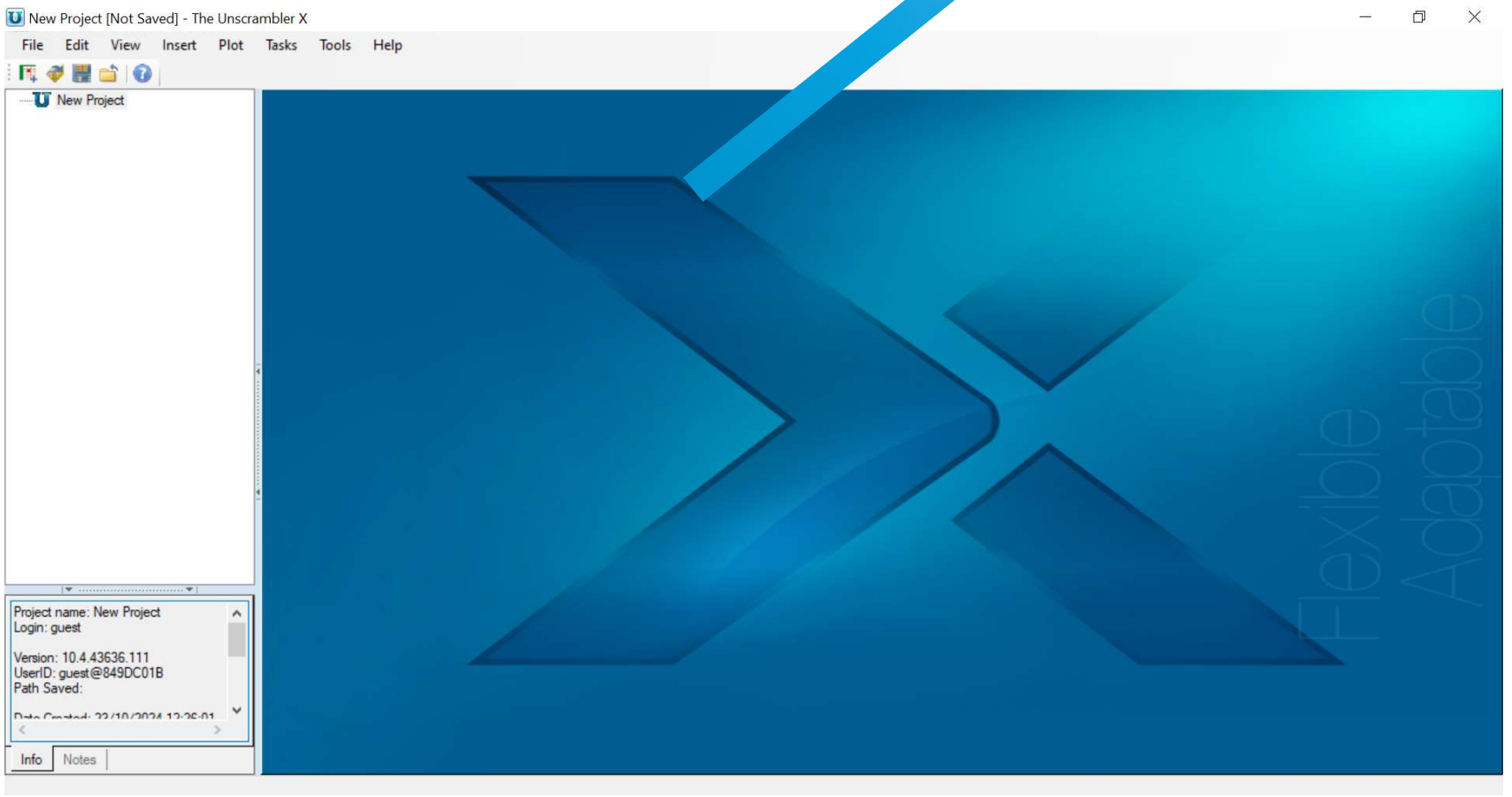


Index

The Unscrambler

Workspace

The **Workspace** occupies the largest area of the application window. It is a **Viewer which displays results either graphically as plots or numerically as tables.**

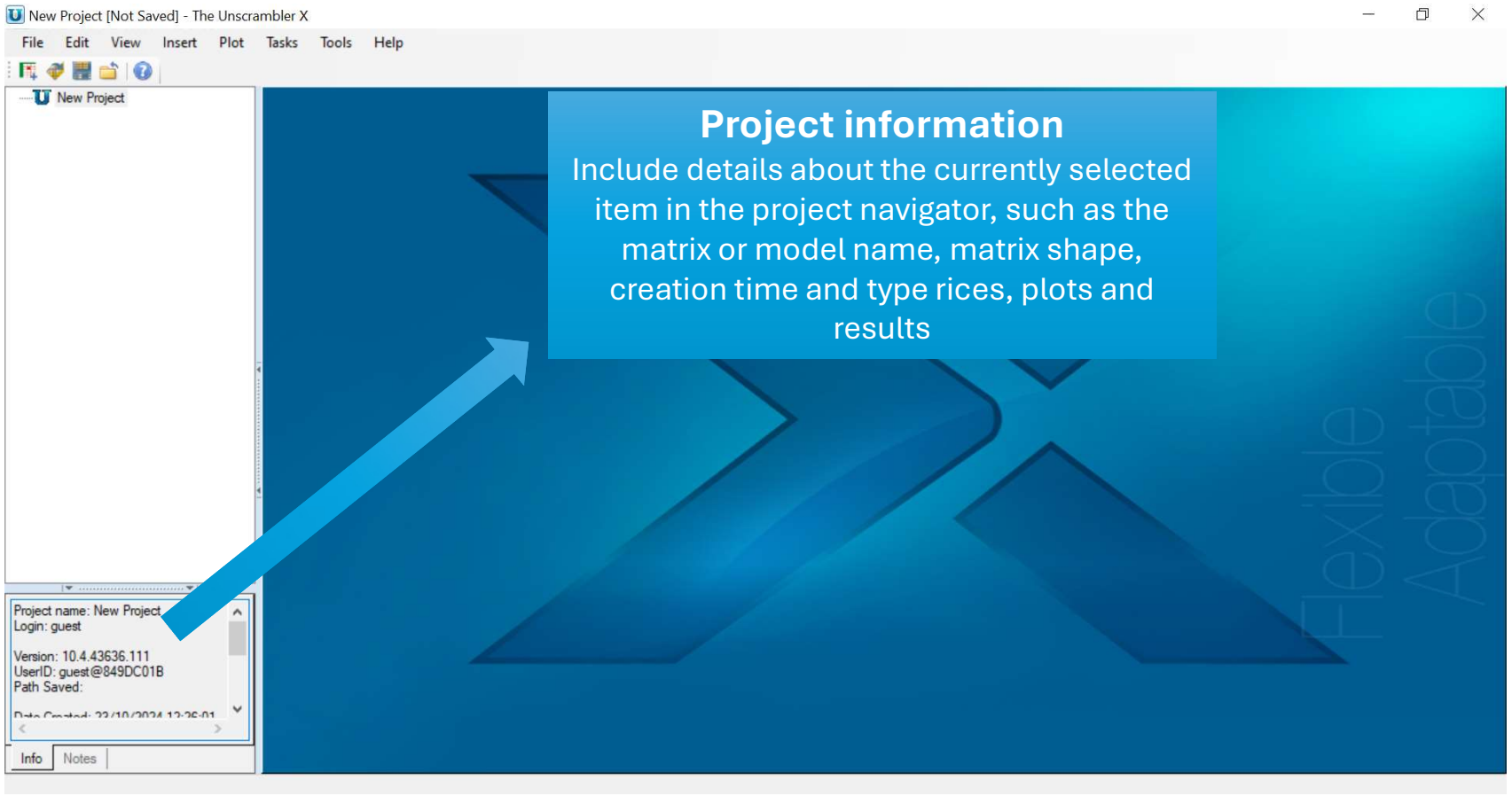


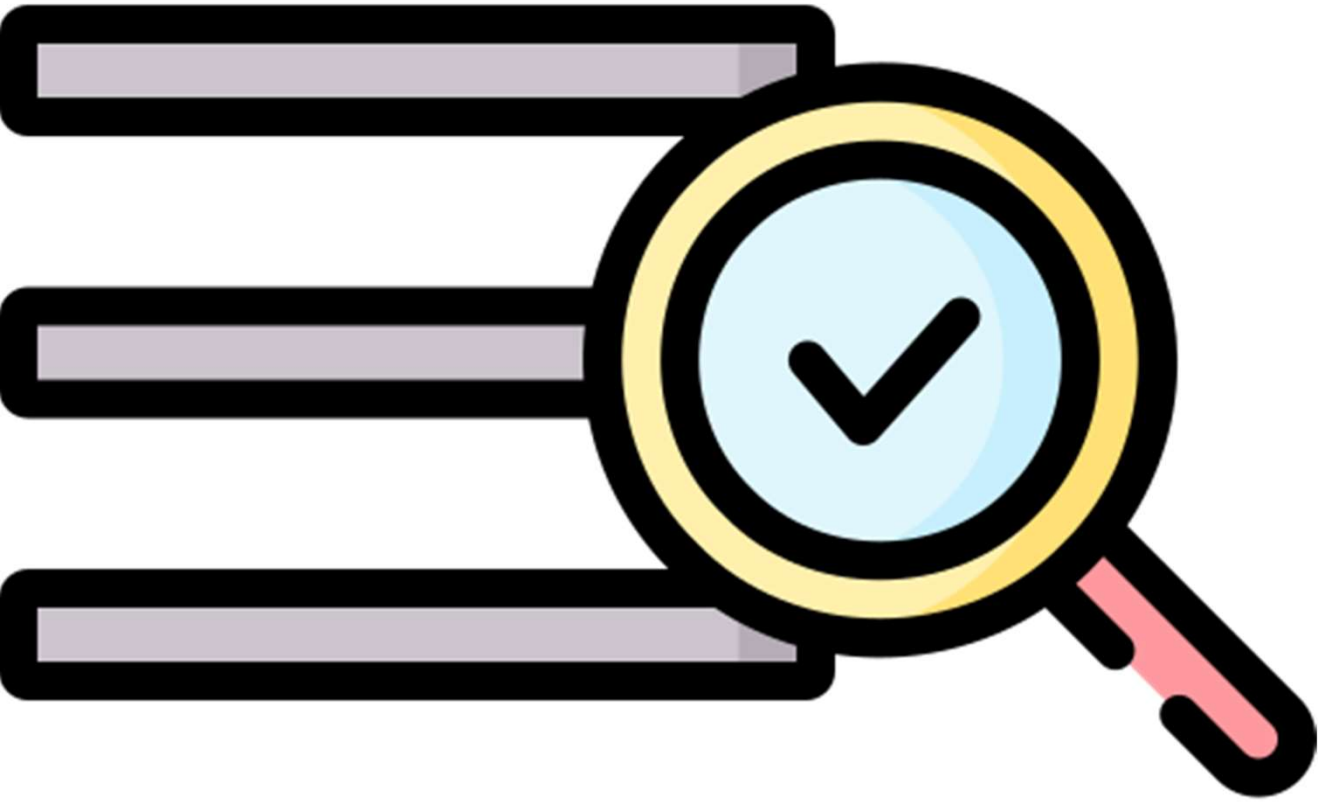
The Unscrambler

Project navigator

Is a **tree-like structure consisting of data matrices and analysis** results along with plots

The Unscrambler





Index

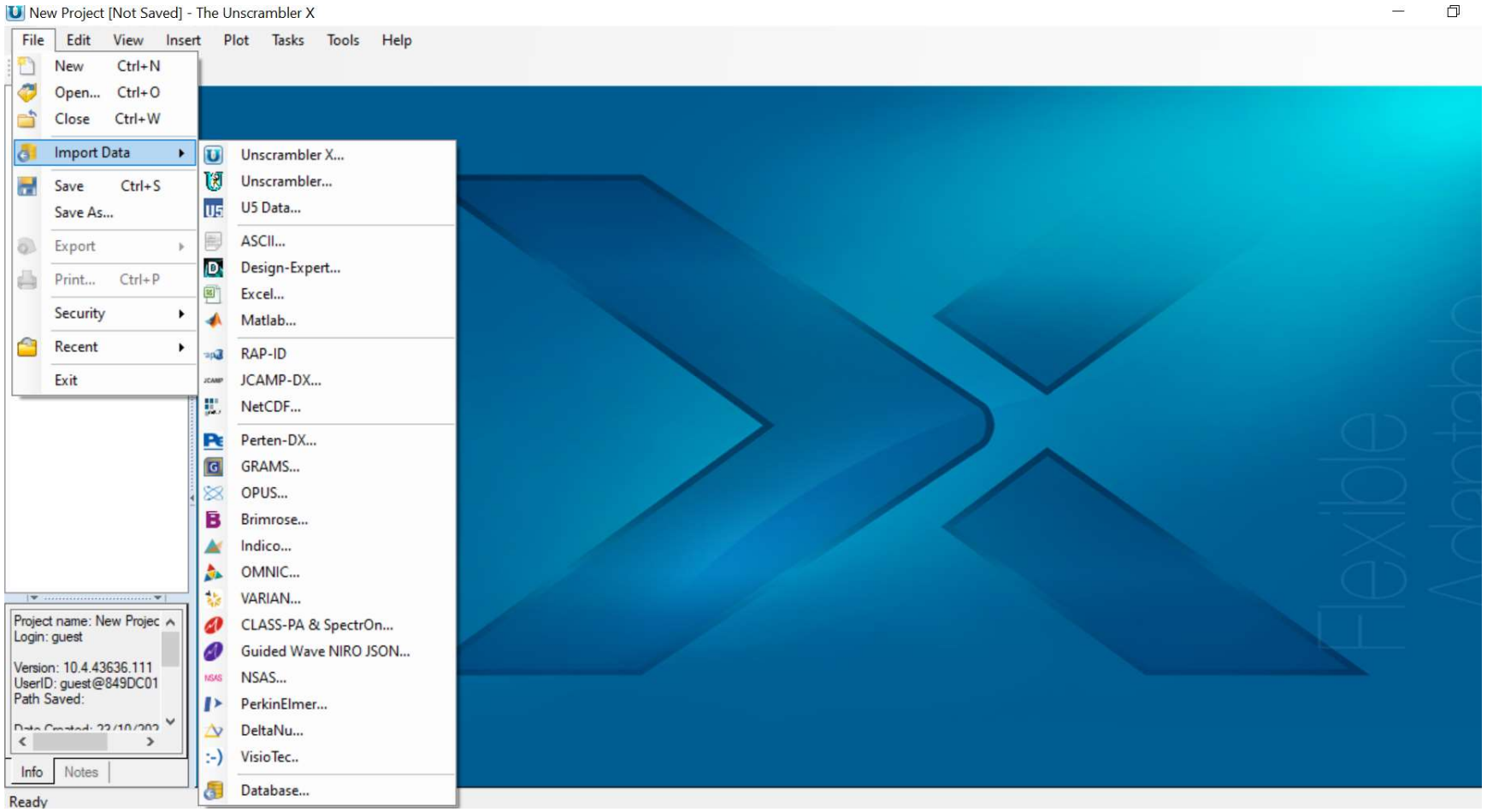
The construction of the matrix

- Data import
- Reduce (average)
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Plot line

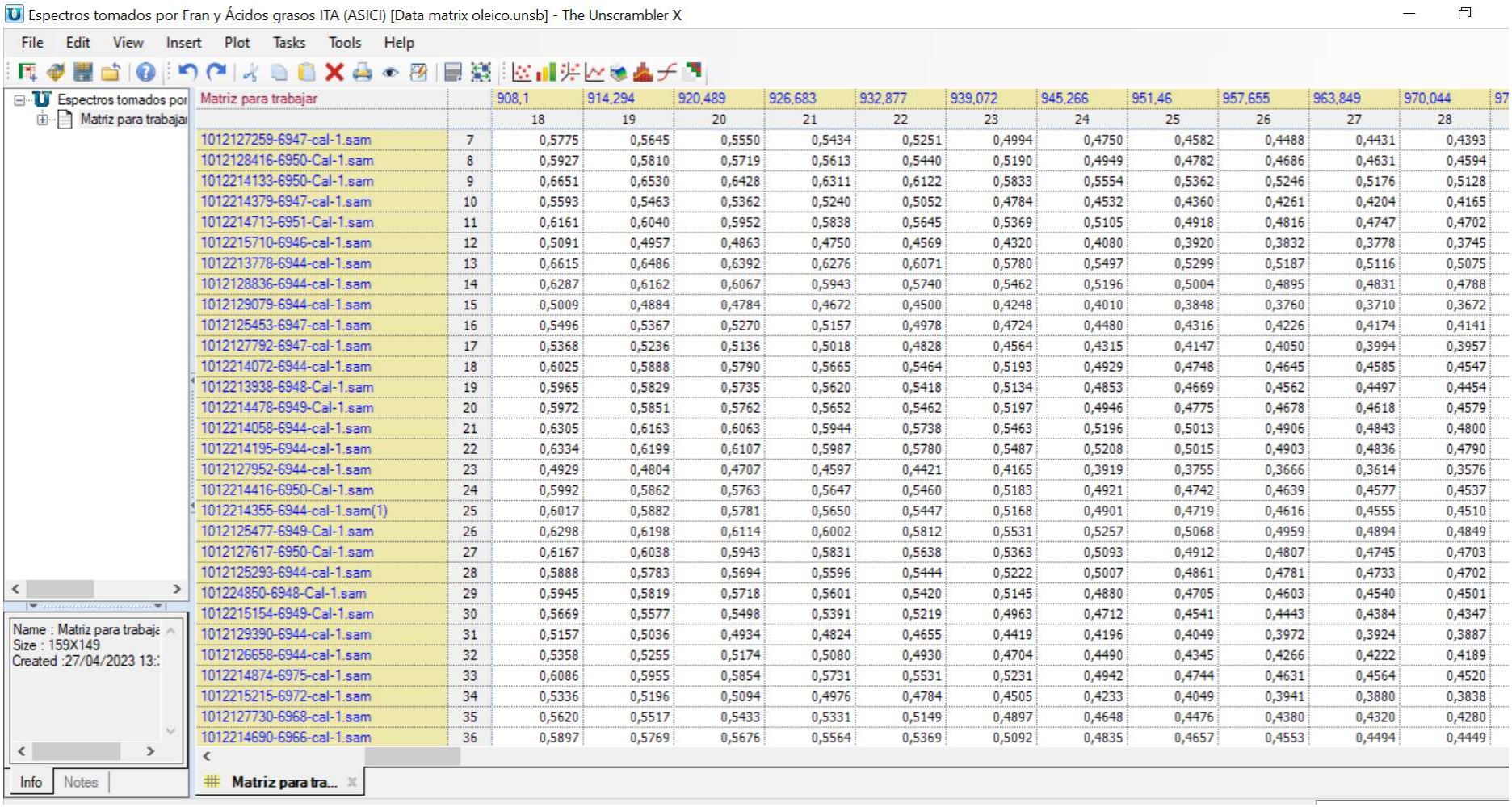
The construction of the matrix

- **Data import**
- Reduce (average)
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Plot line

- Data import


New Project [Not Saved] - The Unscrambler X
File
Edit
View
Insert
Plot
Tasks
Tools
Help
New Ctrl+N
Open... Ctrl+O
Close Ctrl+W
Import Data
Save Ctrl+S
Save As...
Export
Print... Ctrl+P
Security
Recent
Exit
Unscrambler X...
Unscrambler...
U5 Data...
ASCII...
Design-Expert...
Excel...
Matlab...
RAP-ID
JCAMP-DX...
NetCDF...
Perten-DX...
GRAMS...
OPUS...
Brimrose...
Indico...
OMNIC...
VARIAN...
CLASS-PA & SpectrOn...
Guided Wave NIRO JSON...
NSAS...
PerkinElmer...
DeltaNu...
VisioTec...
Database...
Project name: New Projec
Login: guest
Version: 10.4.43636.111
UserID: guest@849DC01
Path Saved:
Info
Notes
Ready
Flexible
Adaptable

- Data import



Matriz para trabajar		908,1	914,294	920,489	926,683	932,877	939,072	945,266	951,46	957,655	963,849	970,044	97
		18	19	20	21	22	23	24	25	26	27	28	
1012127259-6947-cal-1.sam	7	0,5775	0,5645	0,5550	0,5434	0,5251	0,4994	0,4750	0,4582	0,4488	0,4431	0,4393	
1012128416-6950-Cal-1.sam	8	0,5927	0,5810	0,5719	0,5613	0,5440	0,5190	0,4949	0,4782	0,4686	0,4631	0,4594	
1012214133-6950-Cal-1.sam	9	0,6651	0,6530	0,6428	0,6311	0,6122	0,5833	0,5554	0,5362	0,5246	0,5176	0,5128	
1012214379-6947-cal-1.sam	10	0,5593	0,5463	0,5362	0,5240	0,5052	0,4784	0,4532	0,4360	0,4261	0,4204	0,4165	
1012214713-6951-Cal-1.sam	11	0,6161	0,6040	0,5952	0,5838	0,5645	0,5369	0,5105	0,4918	0,4816	0,4747	0,4702	
1012215710-6946-cal-1.sam	12	0,5091	0,4957	0,4863	0,4750	0,4569	0,4320	0,4080	0,3920	0,3832	0,3778	0,3745	
1012213778-6944-cal-1.sam	13	0,6615	0,6486	0,6392	0,6276	0,6071	0,5780	0,5497	0,5299	0,5187	0,5116	0,5075	
1012128836-6944-cal-1.sam	14	0,6287	0,6162	0,6067	0,5943	0,5740	0,5462	0,5196	0,5004	0,4895	0,4831	0,4788	
1012129079-6944-cal-1.sam	15	0,5009	0,4884	0,4784	0,4672	0,4500	0,4248	0,4010	0,3848	0,3760	0,3710	0,3672	
1012125453-6947-cal-1.sam	16	0,5496	0,5367	0,5270	0,5157	0,4978	0,4724	0,4480	0,4316	0,4226	0,4174	0,4141	
1012127792-6947-cal-1.sam	17	0,5368	0,5236	0,5136	0,5018	0,4828	0,4564	0,4315	0,4147	0,4050	0,3994	0,3957	
1012214072-6944-cal-1.sam	18	0,6025	0,5888	0,5790	0,5665	0,5464	0,5193	0,4929	0,4748	0,4645	0,4585	0,4547	
1012213938-6948-Cal-1.sam	19	0,5965	0,5829	0,5735	0,5620	0,5418	0,5134	0,4853	0,4669	0,4562	0,4497	0,4454	
1012214478-6949-Cal-1.sam	20	0,5972	0,5851	0,5762	0,5652	0,5462	0,5197	0,4946	0,4775	0,4678	0,4618	0,4579	
1012214058-6944-cal-1.sam	21	0,6305	0,6163	0,6063	0,5944	0,5738	0,5463	0,5196	0,5013	0,4906	0,4843	0,4800	
1012214195-6944-cal-1.sam	22	0,6334	0,6199	0,6107	0,5987	0,5780	0,5487	0,5208	0,5015	0,4903	0,4836	0,4790	
1012127952-6944-cal-1.sam	23	0,4929	0,4804	0,4707	0,4597	0,4421	0,4165	0,3919	0,3755	0,3666	0,3614	0,3576	
1012214416-6950-Cal-1.sam	24	0,5992	0,5862	0,5763	0,5647	0,5460	0,5183	0,4921	0,4742	0,4639	0,4577	0,4537	
1012214355-6944-cal-1.sam(1)	25	0,6017	0,5882	0,5781	0,5650	0,5447	0,5168	0,4901	0,4719	0,4616	0,4555	0,4510	
1012125477-6949-Cal-1.sam	26	0,6298	0,6198	0,6114	0,6002	0,5812	0,5531	0,5257	0,5068	0,4959	0,4894	0,4849	
1012127617-6950-Cal-1.sam	27	0,6167	0,6038	0,5943	0,5831	0,5638	0,5363	0,5093	0,4912	0,4807	0,4745	0,4703	
1012125293-6944-cal-1.sam	28	0,5888	0,5783	0,5694	0,5596	0,5444	0,5222	0,5007	0,4861	0,4781	0,4733	0,4702	
101224850-6948-Cal-1.sam	29	0,5945	0,5819	0,5718	0,5601	0,5420	0,5145	0,4880	0,4705	0,4603	0,4540	0,4501	
1012215154-6949-Cal-1.sam	30	0,5669	0,5577	0,5498	0,5391	0,5219	0,4963	0,4712	0,4541	0,4443	0,4384	0,4347	
1012129390-6944-cal-1.sam	31	0,5157	0,5036	0,4934	0,4824	0,4655	0,4419	0,4196	0,4049	0,3972	0,3924	0,3887	
1012126658-6944-cal-1.sam	32	0,5358	0,5255	0,5174	0,5080	0,4930	0,4704	0,4490	0,4345	0,4266	0,4222	0,4189	
1012214874-6975-cal-1.sam	33	0,6086	0,5955	0,5854	0,5731	0,5531	0,5231	0,4942	0,4744	0,4631	0,4564	0,4520	
1012215215-6972-cal-1.sam	34	0,5336	0,5196	0,5094	0,4976	0,4784	0,4505	0,4233	0,4049	0,3941	0,3880	0,3838	
1012127730-6968-cal-1.sam	35	0,5620	0,5517	0,5433	0,5331	0,5149	0,4897	0,4648	0,4476	0,4380	0,4320	0,4280	
1012214690-6966-cal-1.sam	36	0,5897	0,5769	0,5676	0,5564	0,5369	0,5092	0,4835	0,4657	0,4553	0,4494	0,4449	

The construction of the matrix

- Data import
- **Reduce (average)**
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Plot line

- Reduce (Average)

It consists of an average of spectra. Since noise is considered as a random instrumental variation, averaging over samples (in case of replicates) may have, depending on the context, the following advantages:

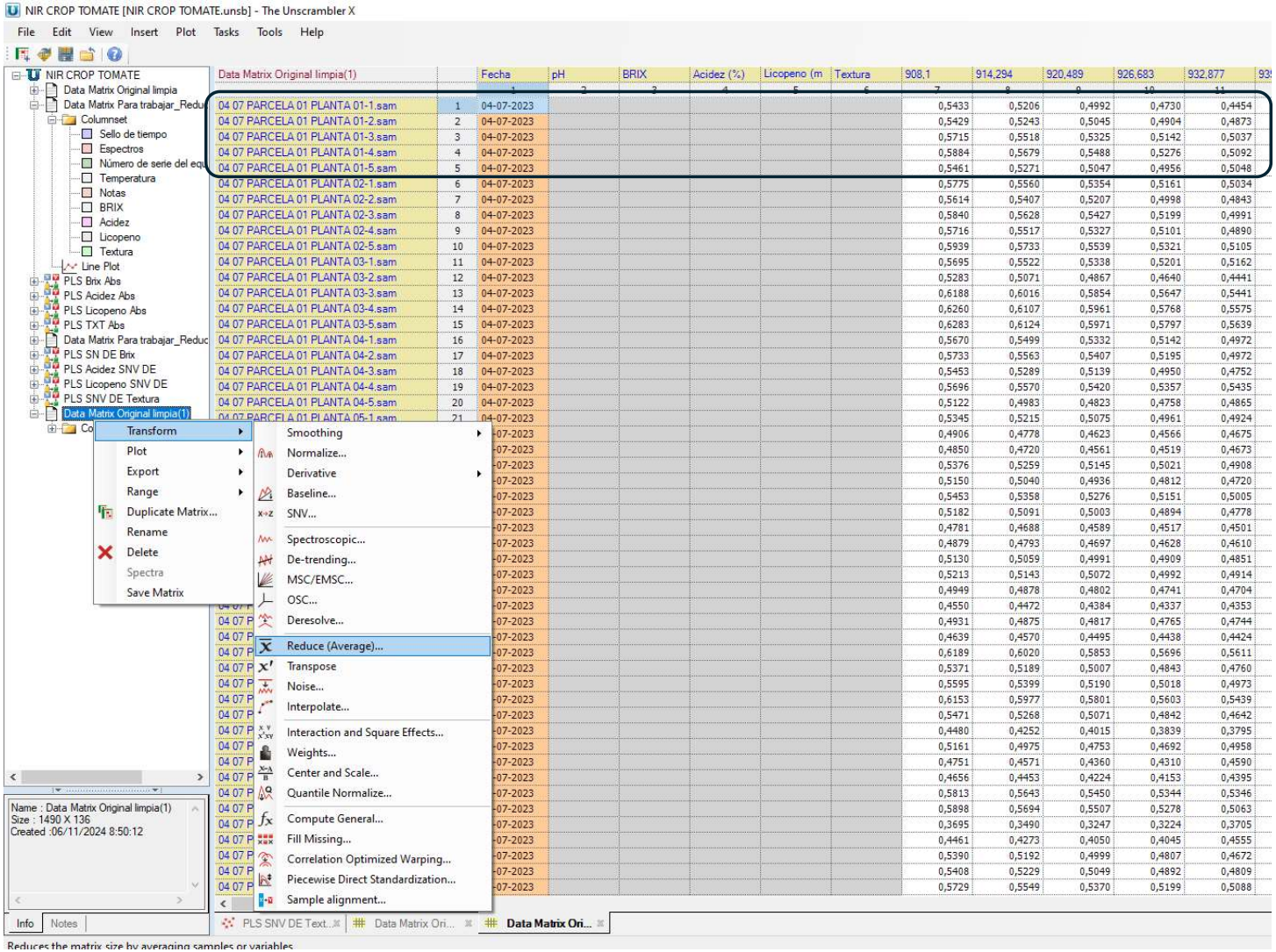
- ❑ **Increase precision**
- ❑ **Get more stable results**
- ❑ **Reduce noise**
- ❑ **Interpret the results more easily**

TE.unsb] - The Unscrambler X

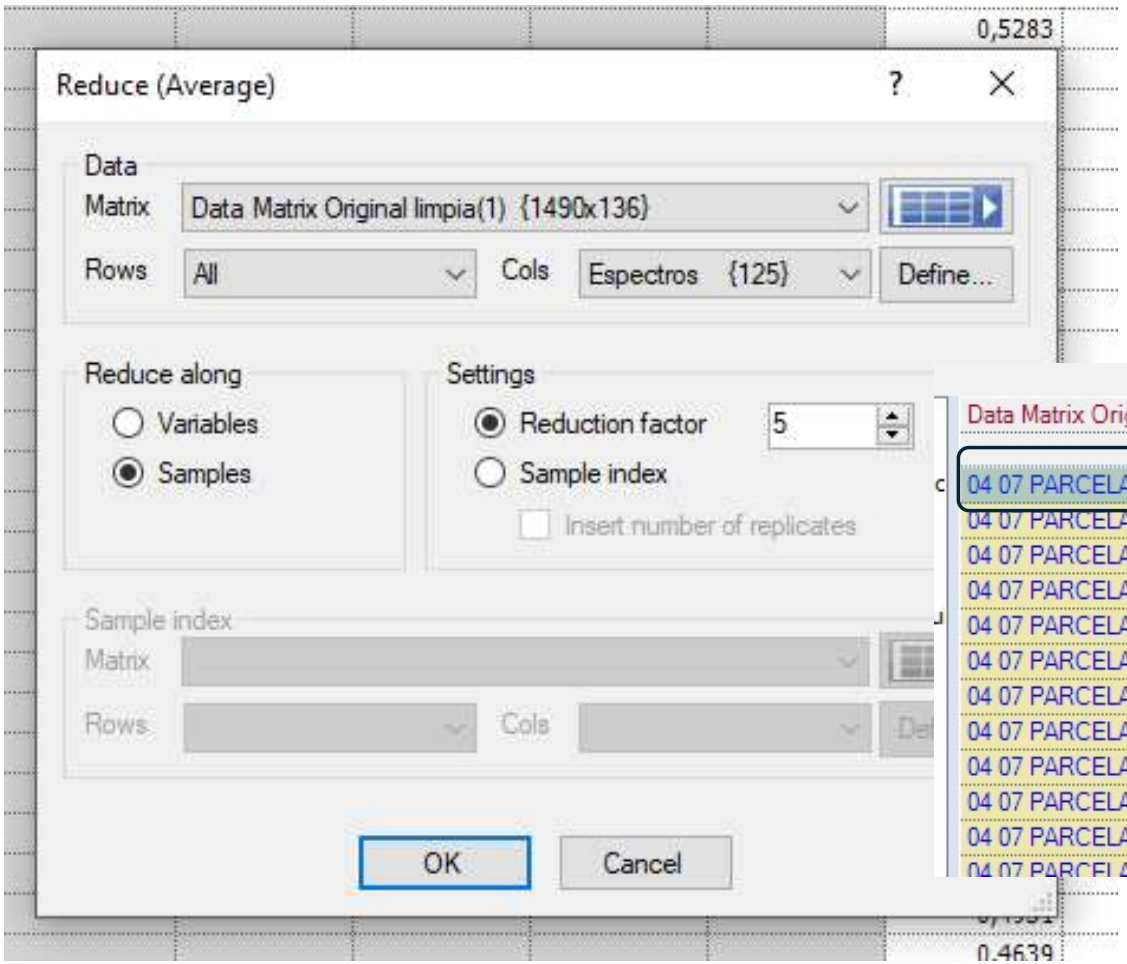
Tasks Tools Help

Data Matrix Original limpia		Fecha								
		1								
04 07 PARCELA 01 PLANTA 01-1.sam	1	04-07-2023								
04 07 PARCELA 01 PLANTA 01-2.sam	2	04-07-2023								
04 07 PARCELA 01 PLANTA 01-3.sam	3	04-07-2023								
04 07 PARCELA 01 PLANTA 01-4.sam	4	04-07-2023								
04 07 PARCELA 01 PLANTA 01-5.sam	5	04-07-2023								
04 07 PARCELA 01 PLANTA 02-1.sam	6	04-07-2023						0,5775	0,5560	0,5354	0,5161
04 07 PARCELA 01 PLANTA 02-2.sam	7	04-07-2023						0,5614	0,5407	0,5207	0,4998
04 07 PARCELA 01 PLANTA 02-3.sam	8	04-07-2023						0,5840	0,5628	0,5427	0,5199
04 07 PARCELA 01 PLANTA 02-4.sam	9	04-07-2023						0,5716	0,5517	0,5327	0,5101
04 07 PARCELA 01 PLANTA 02-5.sam	10	04-07-2023						0,5939	0,5733	0,5539	0,5321
04 07 PARCELA 01 PLANTA 03-1.sam	11	04-07-2023						0,5695	0,5522	0,5338	0,5201
04 07 PARCELA 01 PLANTA 03-2.sam	12	04-07-2023						0,5283	0,5071	0,4867	0,4640
04 07 PARCELA 01 PLANTA 03-3.sam	13	04-07-2023						0,6188	0,6016	0,5854	0,5647
04 07 PARCELA 01 PLANTA 03-4.sam	14	04-07-2023						0,6260	0,6107	0,5961	0,5768
04 07 PARCELA 01 PLANTA 03-5.sam	15	04-07-2023						0,6283	0,6124	0,5971	0,5797
04 07 PARCELA 01 PLANTA 04-1.sam	16	04-07-2023						0,5670	0,5499	0,5332	0,5142
04 07 PARCELA 01 PLANTA 04-2.sam	17	04-07-2023						0,5733	0,5563	0,5407	0,5195
04 07 PARCELA 01 PLANTA 04-3.sam	18	04-07-2023						0,5453	0,5289	0,5139	0,4950
04 07 PARCELA 01 PLANTA 04-4.sam	19	04-07-2023						0,5696	0,5570	0,5420	0,5357
04 07 PARCELA 01 PLANTA 04-5.sam	20	04-07-2023						0,5122	0,4983	0,4823	0,4758
04 07 PARCELA 01 PLANTA 05-1.sam	21	04-07-2023						0,5345	0,5215	0,5075	0,4961
04 07 PARCELA 01 PLANTA 05-2.sam	22	04-07-2023						0,4906	0,4778	0,4623	0,4566
04 07 PARCELA 01 PLANTA 05-3.sam	23	04-07-2023						0,4850	0,4720	0,4561	0,4519
04 07 PARCELA 01 PLANTA 05-4.sam	24	04-07-2023						0,5376	0,5259	0,5145	0,5021
04 07 PARCELA 01 PLANTA 05-5.sam	25	04-07-2023						0,5150	0,5040	0,4936	0,4812

- Reduce (Average)



- Reduce (Average)

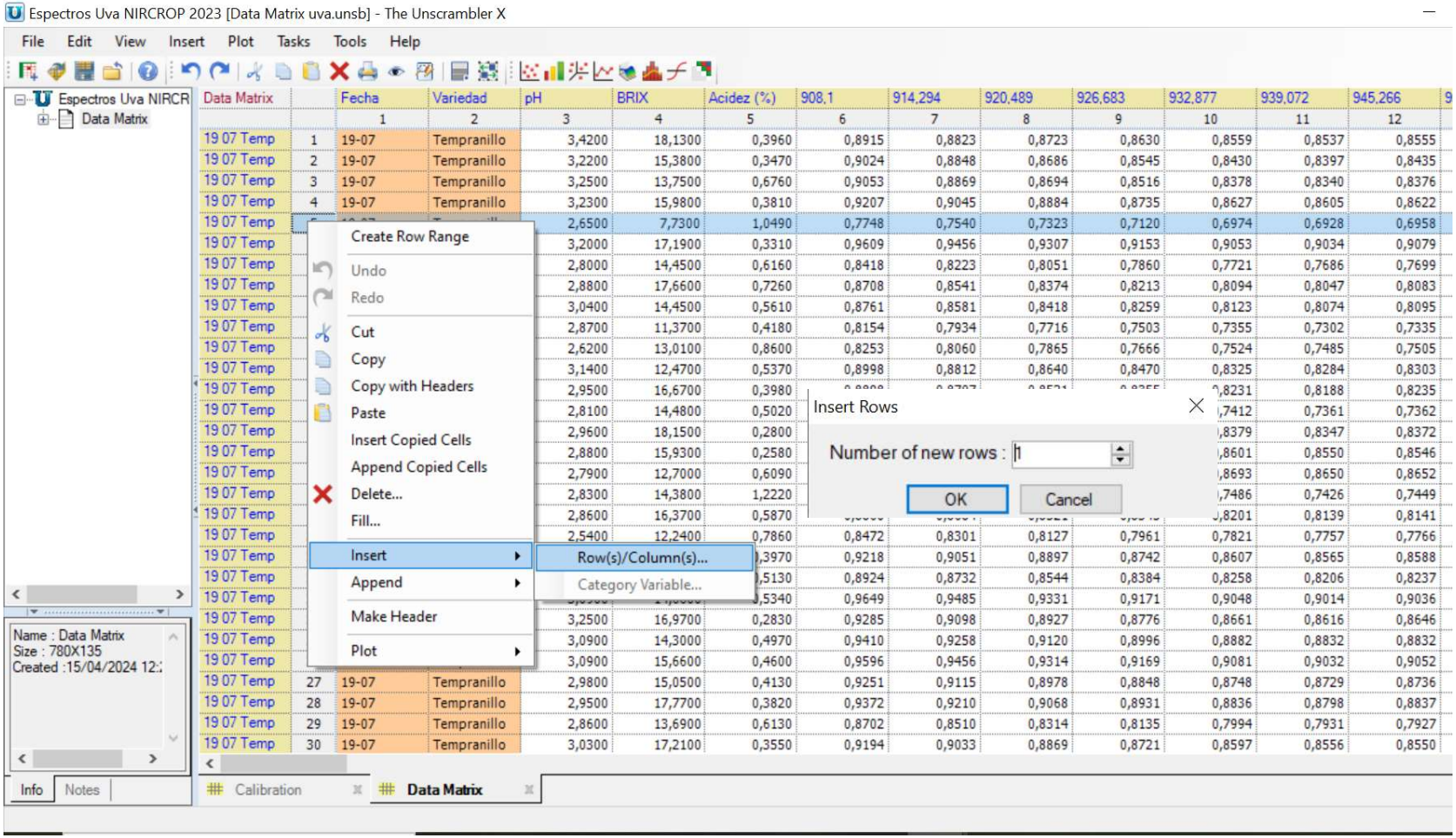


Data Matrix Original limpia(1)_Reduced		908,1	914,294	920,489
		1	2	3
04 07 PARCELA 01 PLANTA 01-3.sam	1	0,5585	0,5383	0,5179
04 07 PARCELA 01 PLANTA 02-3.sam	2	0,5777	0,5569	0,5371
04 07 PARCELA 01 PLANTA 03-3.sam	3	0,5942	0,5768	0,5598
04 07 PARCELA 01 PLANTA 04-3.sam	4	0,5535	0,5381	0,5224
04 07 PARCELA 01 PLANTA 05-3.sam	5	0,5125	0,5003	0,4868
04 07 PARCELA 01 PLANTA 06-3.sam	6	0,5085	0,4998	0,4911
04 07 PARCELA 01 PLANTA 07-3.sam	7	0,4856	0,4788	0,4714
04 07 PARCELA 01 PLANTA 08-3.sam	8	0,5756	0,5570	0,5384
04 07 PARCELA 01 PLANTA 09-3.sam	9	0,4972	0,4779	0,4560
04 07 PARCELA 01 PLANTA 10-3.sam	10	0,4970	0,4776	0,4570
04 07 PARCELA 01 PLANTA 11-3.sam	11	0,5127	0,4938	0,4739
04 07 PARCELA 01 PLANTA 12-3.sam	12	0,4961	0,4765	0,4554

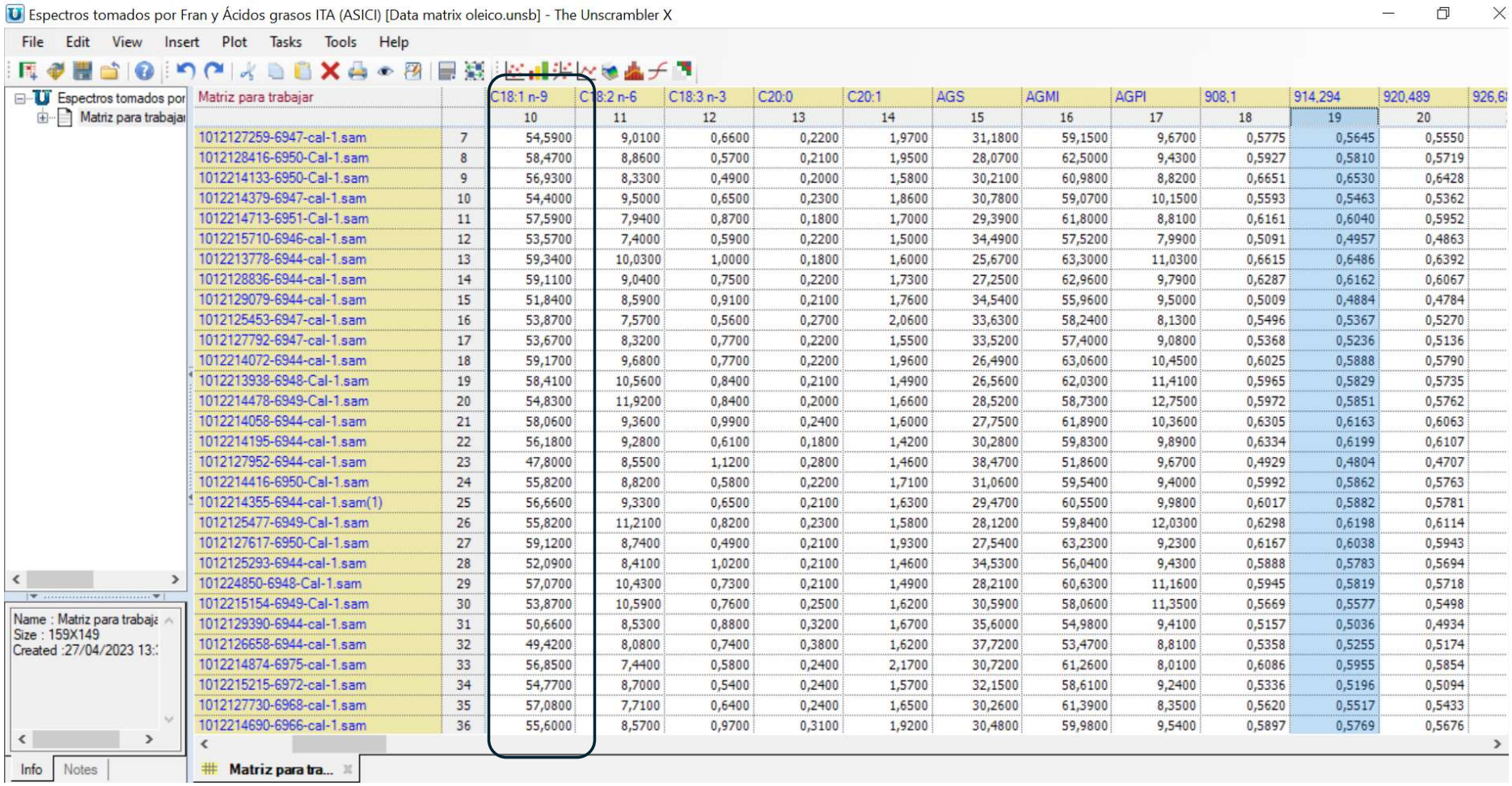
The construction of the matrix

- Data import
- Reduce (average)
- **Insert row/column (to incorporate new samples or physical-chemical variables to predict)**
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Plot line

- Insert row (to incorporate new samples)



- Insert column (to incorporate physical-chemical variables to predict)

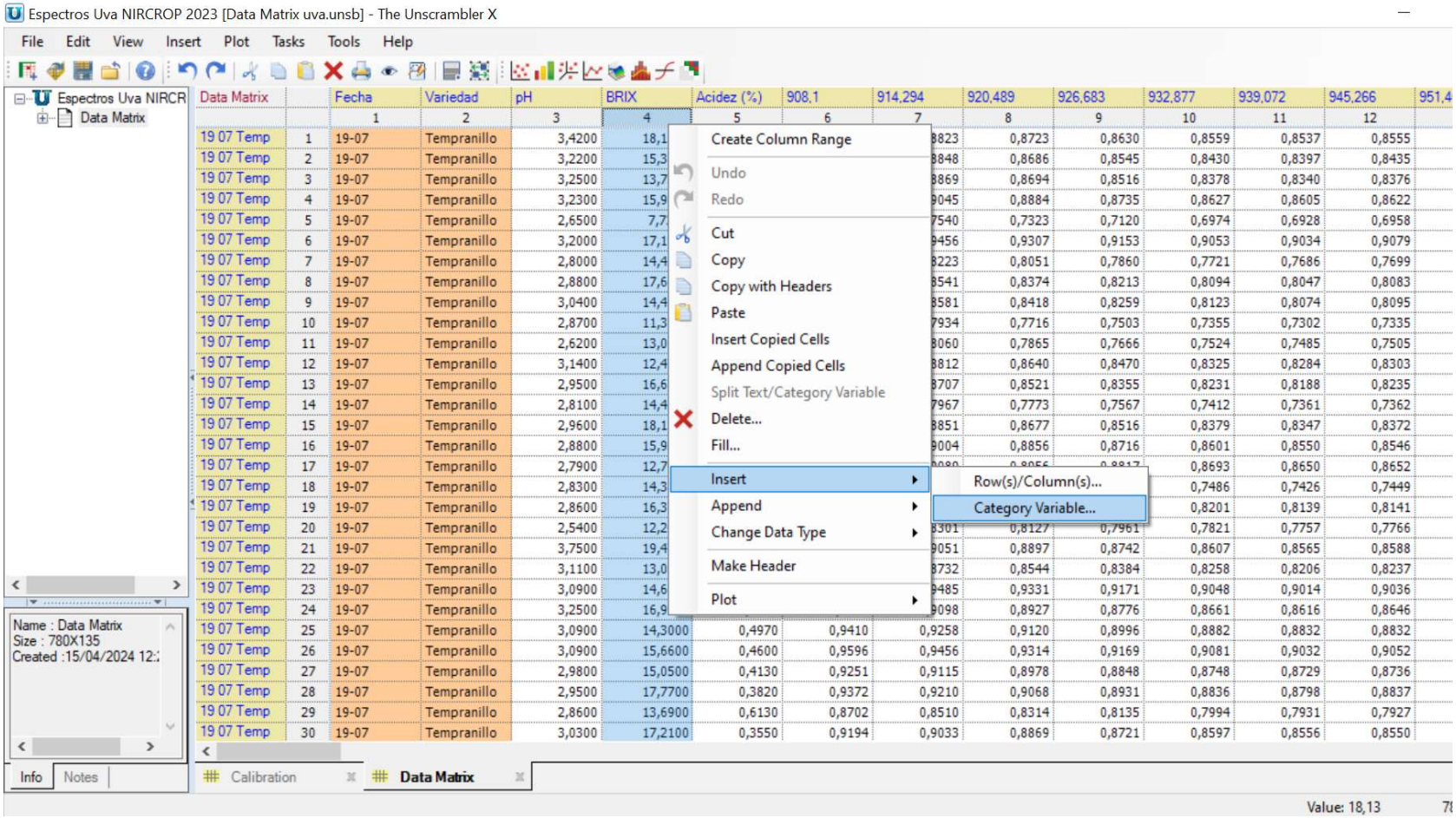


Matriz para trabajar		C18:1 n-9	C18:2 n-6	C18:3 n-3	C20:0	C20:1	AGS	AGMI	AGPI	908,1	914,294	920,489	926,6
		10	11	12	13	14	15	16	17	18	19	20	
1012127259-6947-cal-1.sam	7	54,5900	9,0100	0,6600	0,2200	1,9700	31,1800	59,1500	9,6700	0,5775	0,5645	0,5550	
1012128416-6950-Cal-1.sam	8	58,4700	8,8600	0,5700	0,2100	1,9500	28,0700	62,5000	9,4300	0,5927	0,5810	0,5719	
1012214133-6950-Cal-1.sam	9	56,9300	8,3300	0,4900	0,2000	1,5800	30,2100	60,9800	8,8200	0,6651	0,6530	0,6428	
1012214379-6947-cal-1.sam	10	54,4000	9,5000	0,6500	0,2300	1,8600	30,7800	59,0700	10,1500	0,5593	0,5463	0,5362	
1012214713-6951-Cal-1.sam	11	57,5900	7,9400	0,8700	0,1800	1,7000	29,3900	61,8000	8,8100	0,6161	0,6040	0,5952	
1012215710-6946-cal-1.sam	12	53,5700	7,4000	0,5900	0,2200	1,5000	34,4900	57,5200	7,9900	0,5091	0,4957	0,4863	
1012213778-6944-cal-1.sam	13	59,3400	10,0300	1,0000	0,1800	1,6000	25,6700	63,3000	11,0300	0,6615	0,6486	0,6392	
1012128836-6944-cal-1.sam	14	59,1100	9,0400	0,7500	0,2200	1,7300	27,2500	62,9600	9,7900	0,6287	0,6162	0,6067	
1012129079-6944-cal-1.sam	15	51,8400	8,5900	0,9100	0,2100	1,7600	34,5400	55,9600	9,5000	0,5009	0,4884	0,4784	
1012125453-6947-cal-1.sam	16	53,8700	7,5700	0,5600	0,2700	2,0600	33,6300	58,2400	8,1300	0,5496	0,5367	0,5270	
1012127792-6947-cal-1.sam	17	53,6700	8,3200	0,7700	0,2200	1,5500	33,5200	57,4000	9,0800	0,5368	0,5236	0,5136	
1012214072-6944-cal-1.sam	18	59,1700	9,6800	0,7700	0,2200	1,9600	26,4900	63,0600	10,4500	0,6025	0,5888	0,5790	
1012213938-6948-Cal-1.sam	19	58,4100	10,5600	0,8400	0,2100	1,4900	26,5600	62,0300	11,4100	0,5965	0,5829	0,5735	
1012214478-6949-Cal-1.sam	20	54,8300	11,9200	0,8400	0,2000	1,6600	28,5200	58,7300	12,7500	0,5972	0,5851	0,5762	
1012214058-6944-cal-1.sam	21	58,0600	9,3600	0,9900	0,2400	1,6000	27,7500	61,8900	10,3600	0,6305	0,6163	0,6063	
1012214195-6944-cal-1.sam	22	56,1800	9,2800	0,6100	0,1800	1,4200	30,2800	59,8300	9,8900	0,6334	0,6199	0,6107	
1012127952-6944-cal-1.sam	23	47,8000	8,5500	1,1200	0,2800	1,4600	38,4700	51,8600	9,6700	0,4929	0,4804	0,4707	
1012214416-6950-Cal-1.sam	24	55,8200	8,8200	0,5800	0,2200	1,7100	31,0600	59,5400	9,4000	0,5992	0,5862	0,5763	
1012214355-6944-cal-1.sam(1)	25	56,6600	9,3300	0,6500	0,2100	1,6300	29,4700	60,5500	9,9800	0,6017	0,5882	0,5781	
1012125477-6949-Cal-1.sam	26	55,8200	11,2100	0,8200	0,2300	1,5800	28,1200	59,8400	12,0300	0,6298	0,6198	0,6114	
1012127617-6950-Cal-1.sam	27	59,1200	8,7400	0,4900	0,2100	1,9300	27,5400	63,2300	9,2300	0,6167	0,6038	0,5943	
1012125293-6944-cal-1.sam	28	52,0900	8,4100	1,0200	0,2100	1,4600	34,5300	56,0400	9,4300	0,5888	0,5783	0,5694	
101224850-6948-Cal-1.sam	29	57,0700	10,4300	0,7300	0,2100	1,4900	28,2100	60,6300	11,1600	0,5945	0,5819	0,5718	
1012215154-6949-Cal-1.sam	30	53,8700	10,5900	0,7600	0,2500	1,6200	30,5900	58,0600	11,3500	0,5669	0,5577	0,5498	
1012129390-6944-cal-1.sam	31	50,6600	8,5300	0,8800	0,3200	1,6700	35,6000	54,9800	9,4100	0,5157	0,5036	0,4934	
1012126658-6944-cal-1.sam	32	49,4200	8,0800	0,7400	0,3800	1,6200	37,7200	53,4700	8,8100	0,5358	0,5255	0,5174	
1012214874-6975-cal-1.sam	33	56,8500	7,4400	0,5800	0,2400	2,1700	30,7200	61,2600	8,0100	0,6086	0,5955	0,5854	
1012215215-6972-cal-1.sam	34	54,7700	8,7000	0,5400	0,2400	1,5700	32,1500	58,6100	9,2400	0,5336	0,5196	0,5094	
1012127730-6968-cal-1.sam	35	57,0800	7,7100	0,6400	0,2400	1,6500	30,2600	61,3900	8,3500	0,5620	0,5517	0,5433	
1012214690-6966-cal-1.sam	36	55,6000	8,5700	0,9700	0,3100	1,9200	30,4800	59,9800	9,5400	0,5897	0,5769	0,5676	

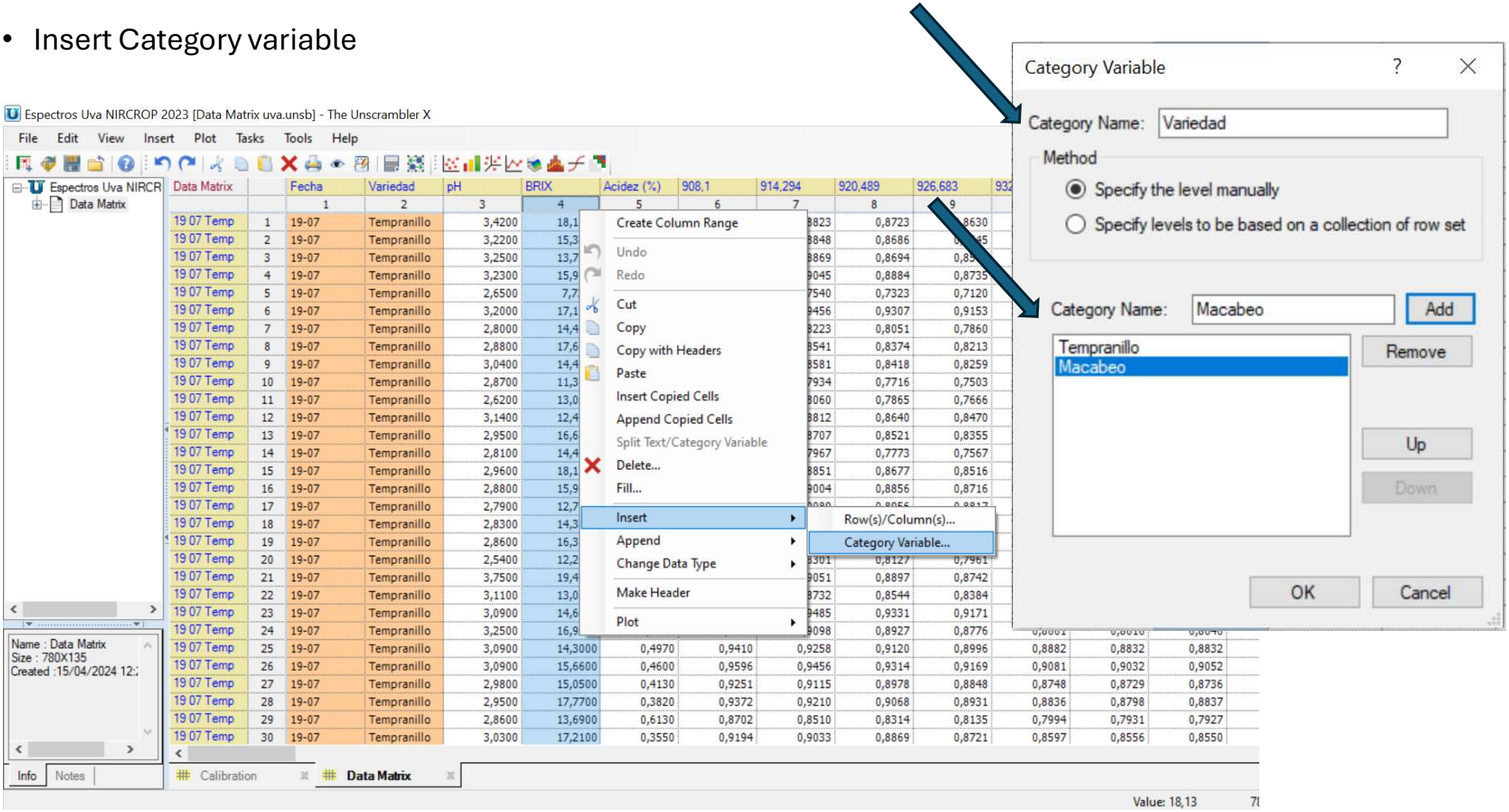
The construction of the matrix

- Data import
- Reduce (average)
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- **Insert Category variable**
- Define range for the purpose of grouping rows and columns
- Plot line

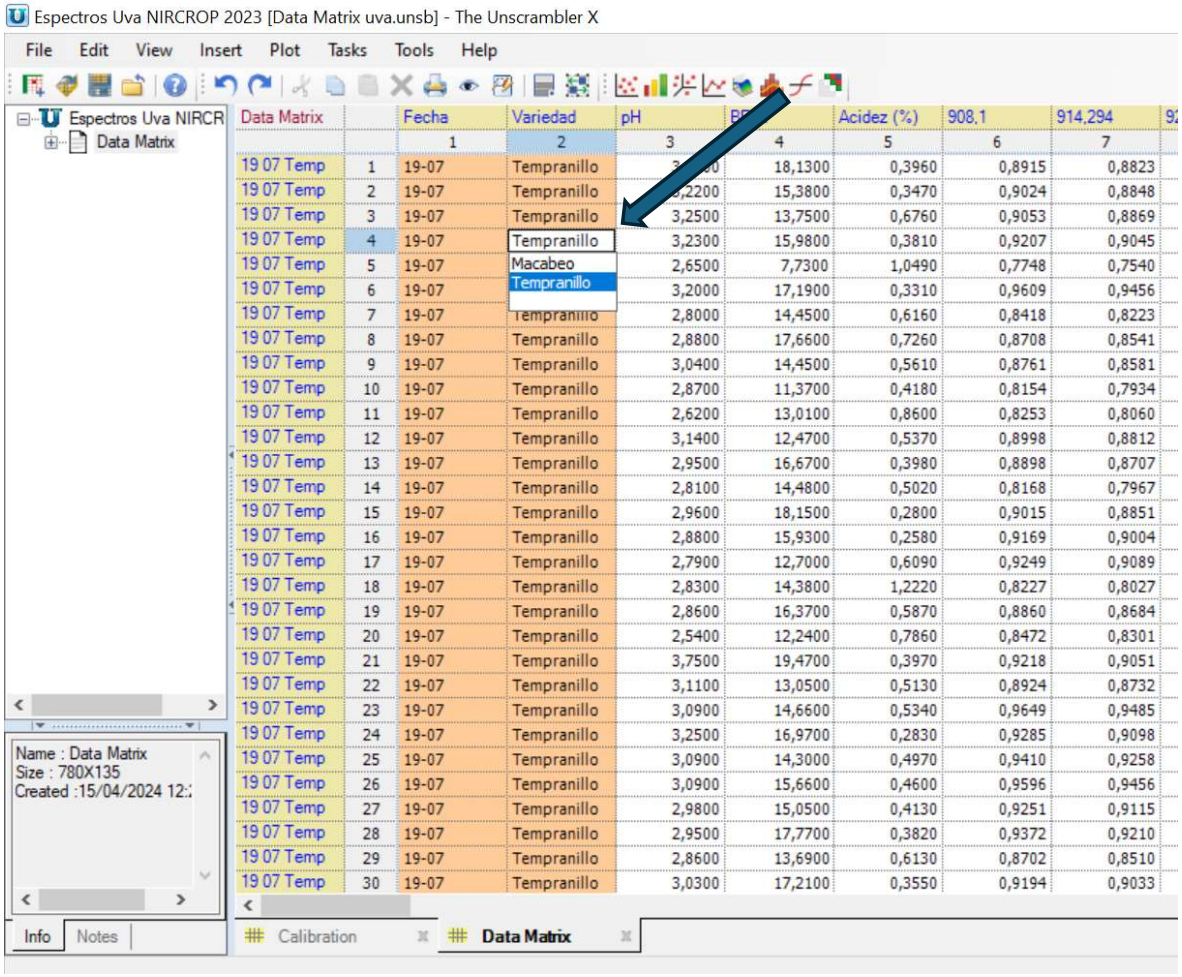
- Insert Category variable



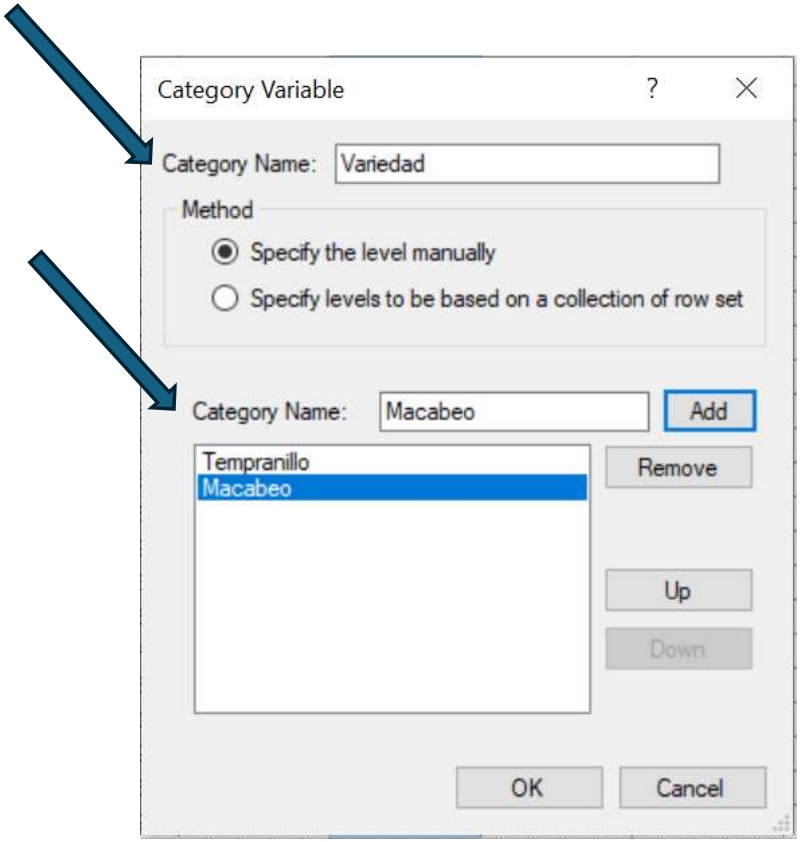
- Insert Category variable



- Insert Category variable



Data Matrix		Fecha	Variedad	pH		Acidez (%)	908,1	914,294	
		1	2	3	4	5	6	7	
19 07 Temp	1	19-07	Tempranillo		18,1300	0,3960	0,8915	0,8823	
19 07 Temp	2	19-07	Tempranillo		15,3800	0,3470	0,9024	0,8848	
19 07 Temp	3	19-07	Tempranillo	3,2500	13,7500	0,6760	0,9053	0,8869	
19 07 Temp	4	19-07	Tempranillo	3,2300	15,9800	0,3810	0,9207	0,9045	
19 07 Temp	5	19-07	Macabeo	2,6500	7,7300	1,0490	0,7748	0,7540	
19 07 Temp	6	19-07	Tempranillo	3,2000	17,1900	0,3310	0,9609	0,9456	
19 07 Temp	7	19-07	Tempranillo	2,8000	14,4500	0,6160	0,8418	0,8223	
19 07 Temp	8	19-07	Tempranillo	2,8800	17,6600	0,7260	0,8708	0,8541	
19 07 Temp	9	19-07	Tempranillo	3,0400	14,4500	0,5610	0,8761	0,8581	
19 07 Temp	10	19-07	Tempranillo	2,8700	11,3700	0,4180	0,8154	0,7934	
19 07 Temp	11	19-07	Tempranillo	2,6200	13,0100	0,8600	0,8253	0,8060	
19 07 Temp	12	19-07	Tempranillo	3,1400	12,4700	0,5370	0,8998	0,8812	
19 07 Temp	13	19-07	Tempranillo	2,9500	16,6700	0,3980	0,8898	0,8707	
19 07 Temp	14	19-07	Tempranillo	2,8100	14,4800	0,5020	0,8168	0,7967	
19 07 Temp	15	19-07	Tempranillo	2,9600	18,1500	0,2800	0,9015	0,8851	
19 07 Temp	16	19-07	Tempranillo	2,8800	15,9300	0,2580	0,9169	0,9004	
19 07 Temp	17	19-07	Tempranillo	2,7900	12,7000	0,6090	0,9249	0,9089	
19 07 Temp	18	19-07	Tempranillo	2,8300	14,3800	1,2220	0,8227	0,8027	
19 07 Temp	19	19-07	Tempranillo	2,8600	16,3700	0,5870	0,8860	0,8684	
19 07 Temp	20	19-07	Tempranillo	2,5400	12,2400	0,7860	0,8472	0,8301	
19 07 Temp	21	19-07	Tempranillo	3,7500	19,4700	0,3970	0,9218	0,9051	
19 07 Temp	22	19-07	Tempranillo	3,1100	13,0500	0,5130	0,8924	0,8732	
19 07 Temp	23	19-07	Tempranillo	3,0900	14,6600	0,5340	0,9649	0,9485	
19 07 Temp	24	19-07	Tempranillo	3,2500	16,9700	0,2830	0,9285	0,9098	
19 07 Temp	25	19-07	Tempranillo	3,0900	14,3000	0,4970	0,9410	0,9258	
19 07 Temp	26	19-07	Tempranillo	3,0900	15,6600	0,4600	0,9596	0,9456	
19 07 Temp	27	19-07	Tempranillo	2,9800	15,0500	0,4130	0,9251	0,9115	
19 07 Temp	28	19-07	Tempranillo	2,9500	17,7700	0,3820	0,9372	0,9210	
19 07 Temp	29	19-07	Tempranillo	2,8600	13,6900	0,6130	0,8702	0,8510	
19 07 Temp	30	19-07	Tempranillo	3,0300	17,2100	0,3550	0,9194	0,9033	

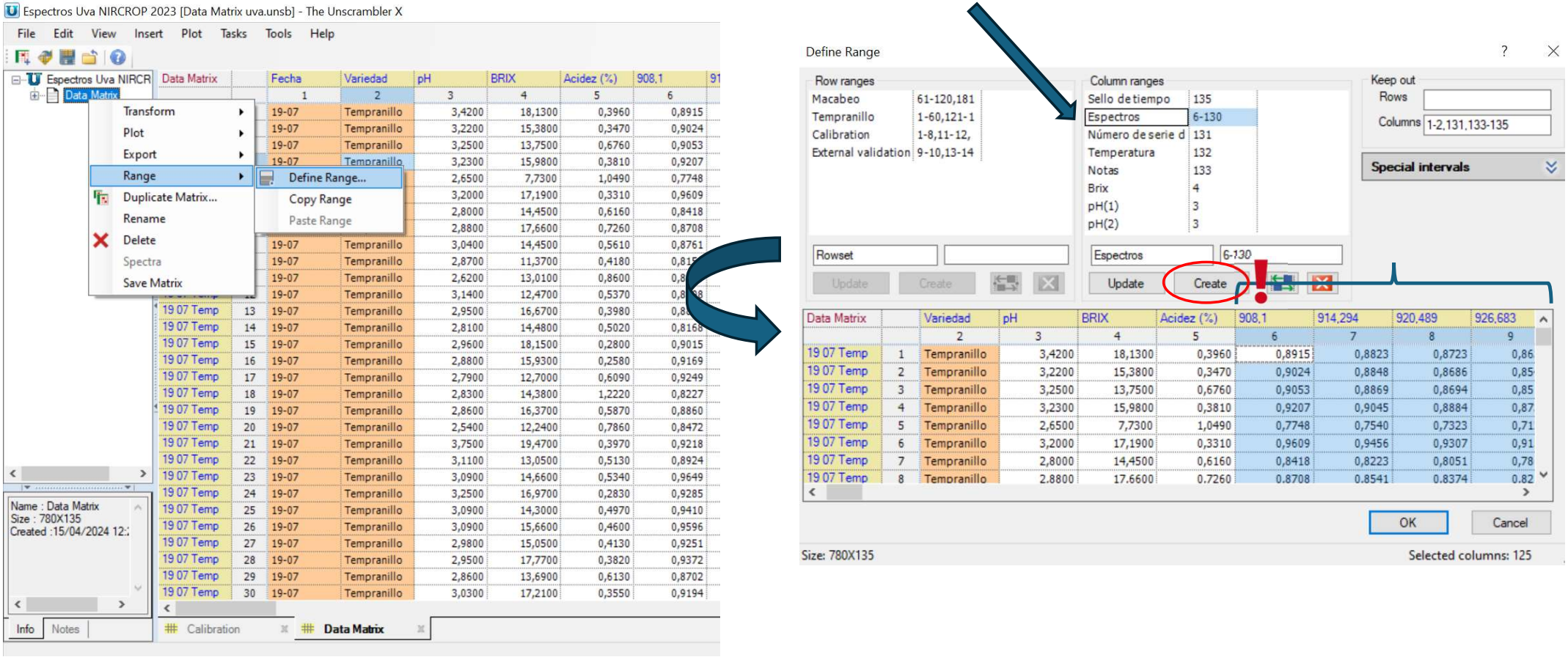


Right side/fill

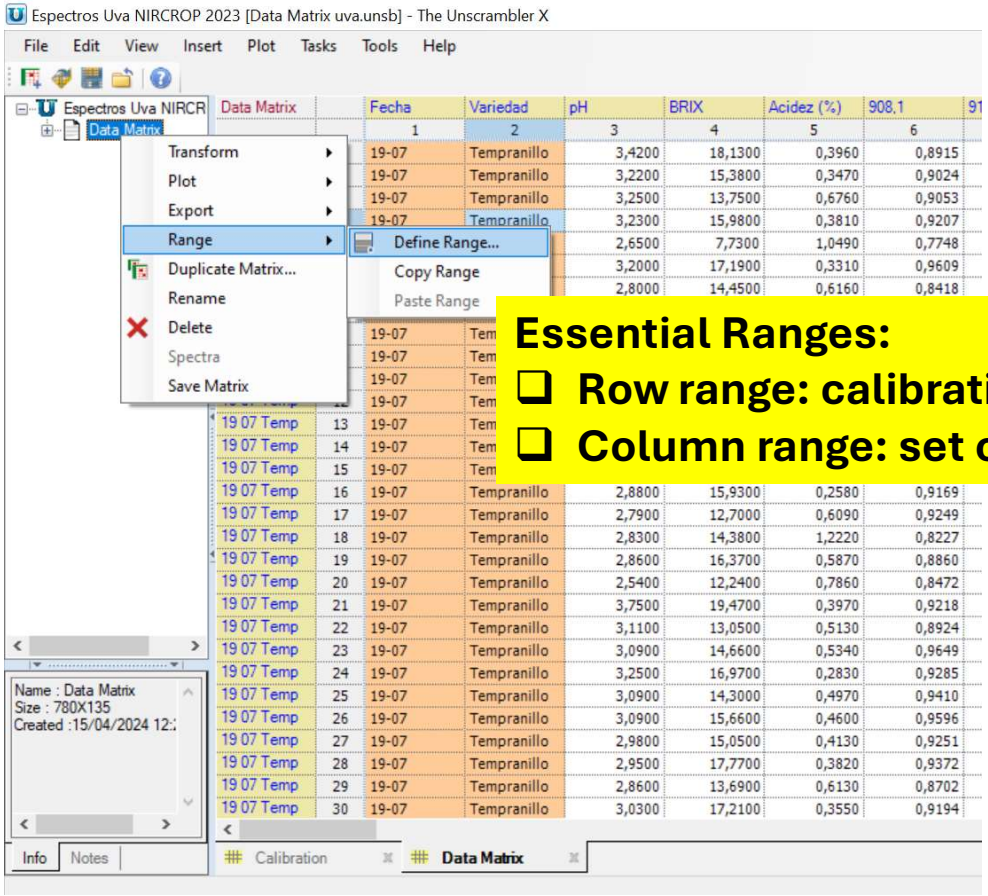
The construction of the matrix

- Data import
- Reduce (average)
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- **Define range for the purpose of grouping rows and columns**
- Plot line

- Define range for the purpose of grouping rows and columns

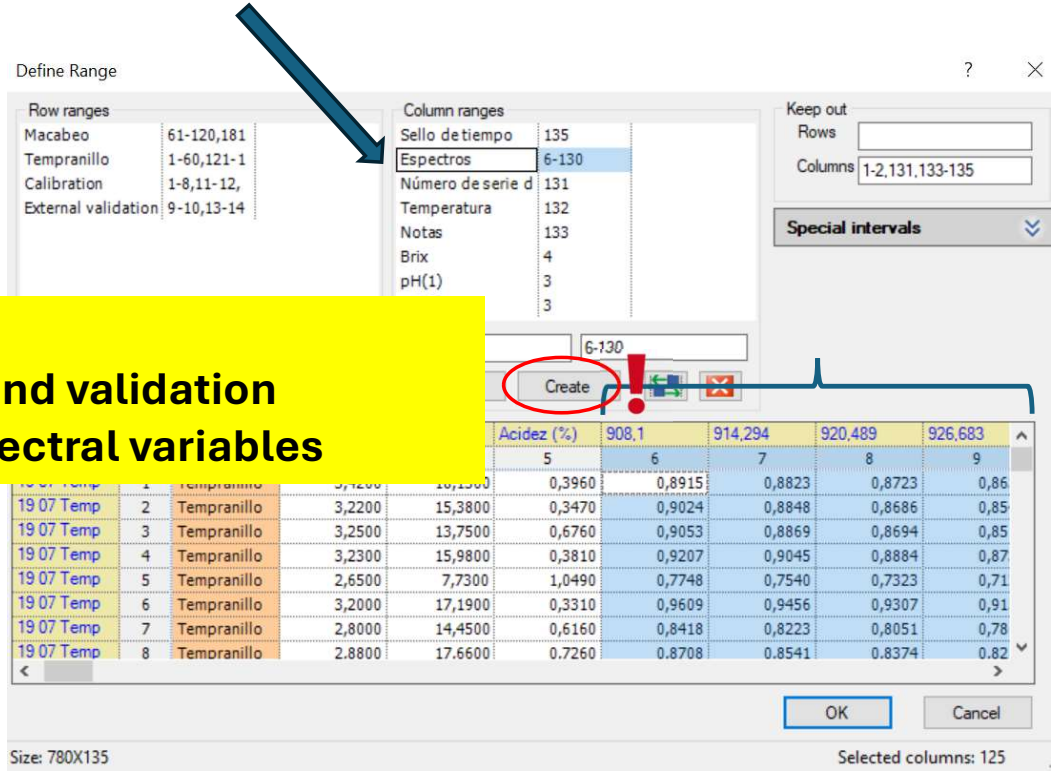


- Define range for the purpose of grouping rows and columns



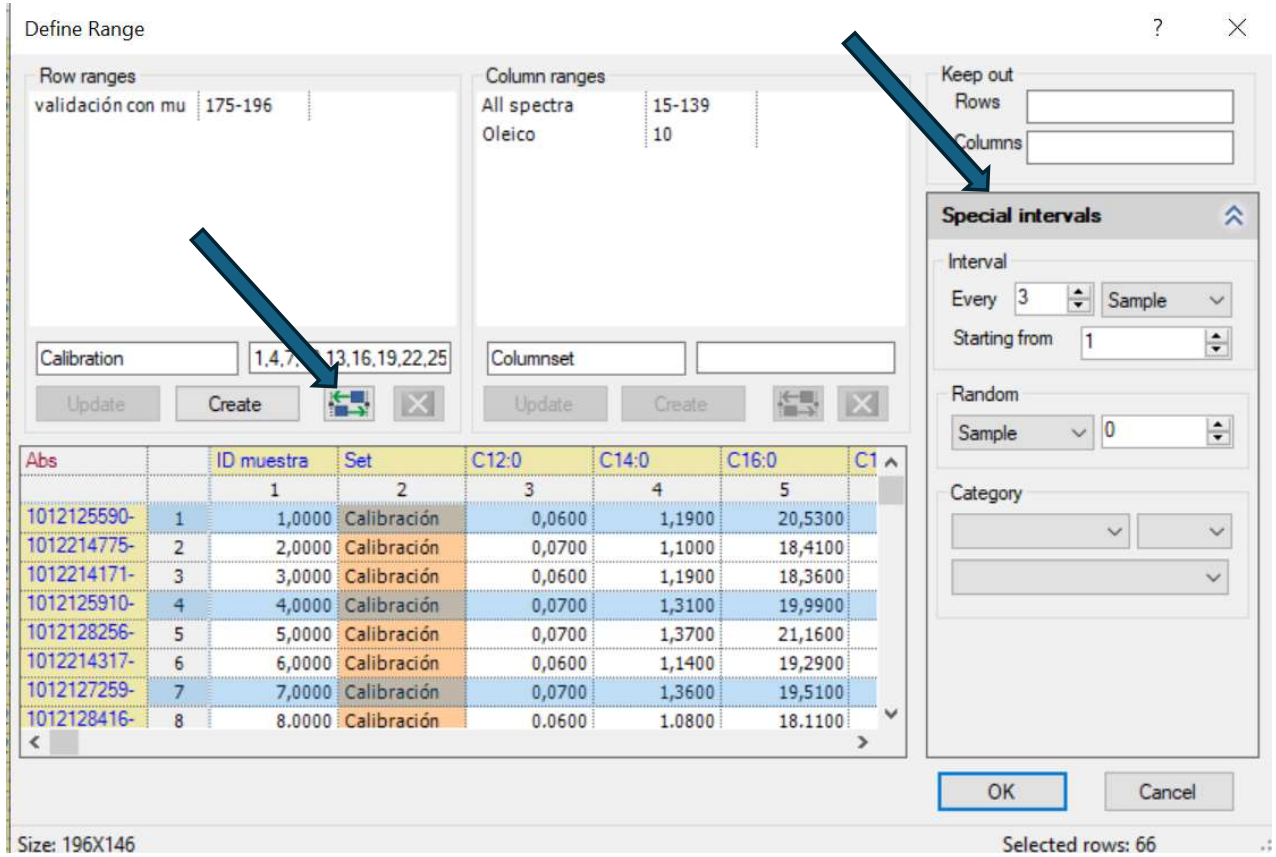
Essential Ranges:

- ❑ **Row range: calibration and validation**
- ❑ **Column range: set of spectral variables**



- Define range for the purpose of grouping rows and columns

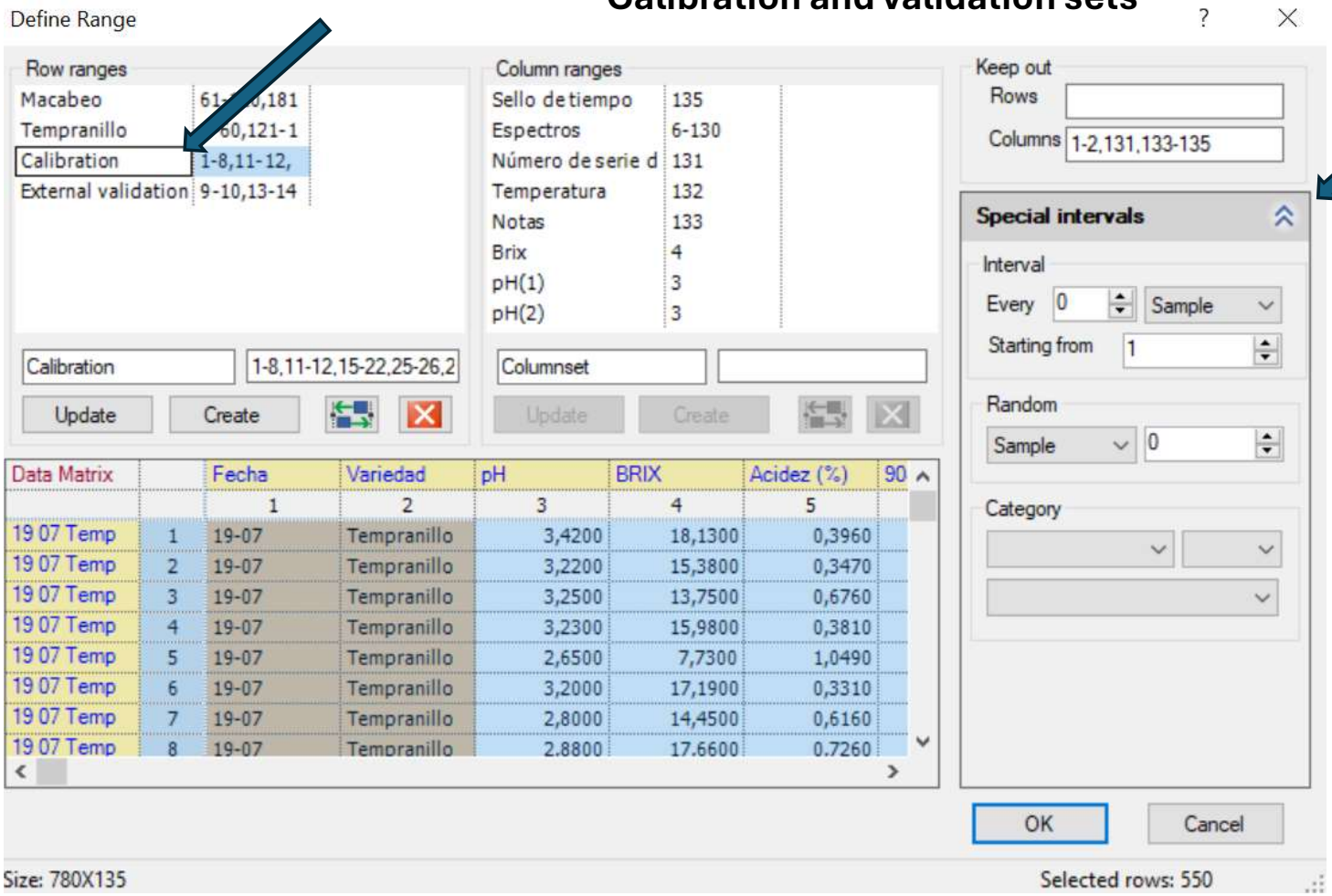
Calibration and validation sets



Abs		ID muestra	Set	C12:0	C14:0	C16:0	C1
		1	2	3	4	5	
1012125590-	1	1,0000	Calibración	0,0600	1,1900	20,5300	
1012214775-	2	2,0000	Calibración	0,0700	1,1000	18,4100	
1012214171-	3	3,0000	Calibración	0,0600	1,1900	18,3600	
1012125910-	4	4,0000	Calibración	0,0700	1,3100	19,9900	
1012128256-	5	5,0000	Calibración	0,0700	1,3700	21,1600	
1012214317-	6	6,0000	Calibración	0,0600	1,1400	19,2900	
1012127259-	7	7,0000	Calibración	0,0700	1,3600	19,5100	
1012128416-	8	8.0000	Calibración	0.0600	1.0800	18.1100	

- Define range for the purpose of grouping rows and columns

Calibration and validation sets



Data Matrix		Fecha	Variedad	pH	BRIX	Acidez (%)	90
		1	2	3	4	5	
19 07 Temp	1	19-07	Tempranillo	3,4200	18,1300	0,3960	
19 07 Temp	2	19-07	Tempranillo	3,2200	15,3800	0,3470	
19 07 Temp	3	19-07	Tempranillo	3,2500	13,7500	0,6760	
19 07 Temp	4	19-07	Tempranillo	3,2300	15,9800	0,3810	
19 07 Temp	5	19-07	Tempranillo	2,6500	7,7300	1,0490	
19 07 Temp	6	19-07	Tempranillo	3,2000	17,1900	0,3310	
19 07 Temp	7	19-07	Tempranillo	2,8000	14,4500	0,6160	
19 07 Temp	8	19-07	Tempranillo	2,8800	17,6600	0,7260	

It is also possible to define calibration and validation sets by means of a category variable

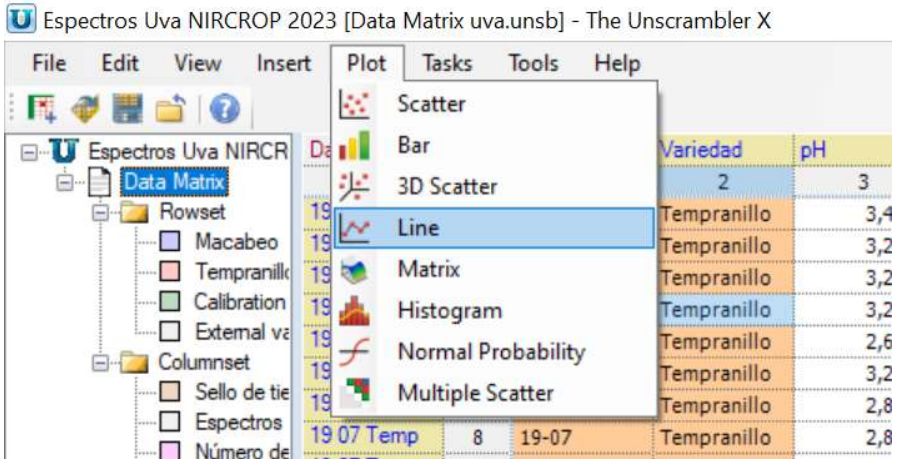
The construction of the matrix

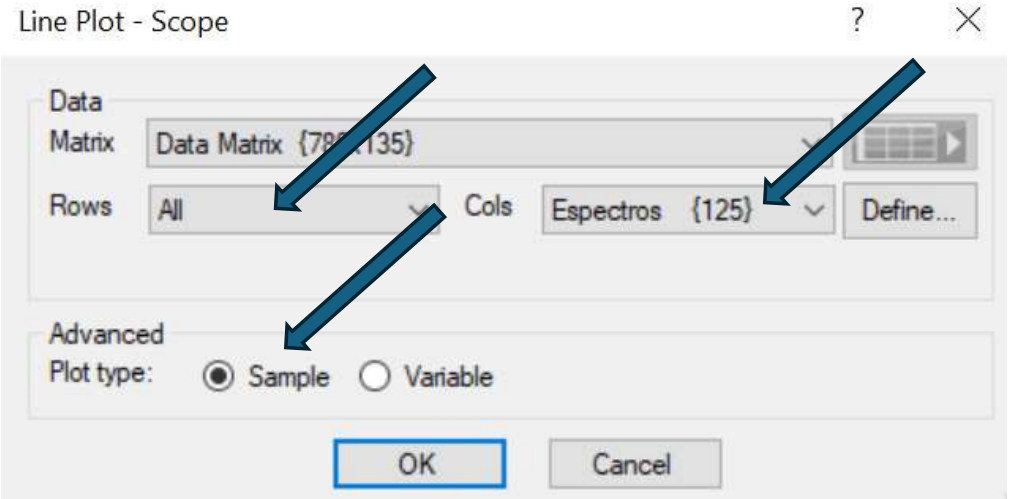
- Data import
- Reduce (average)
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)
- **Plot line**

- Plot line

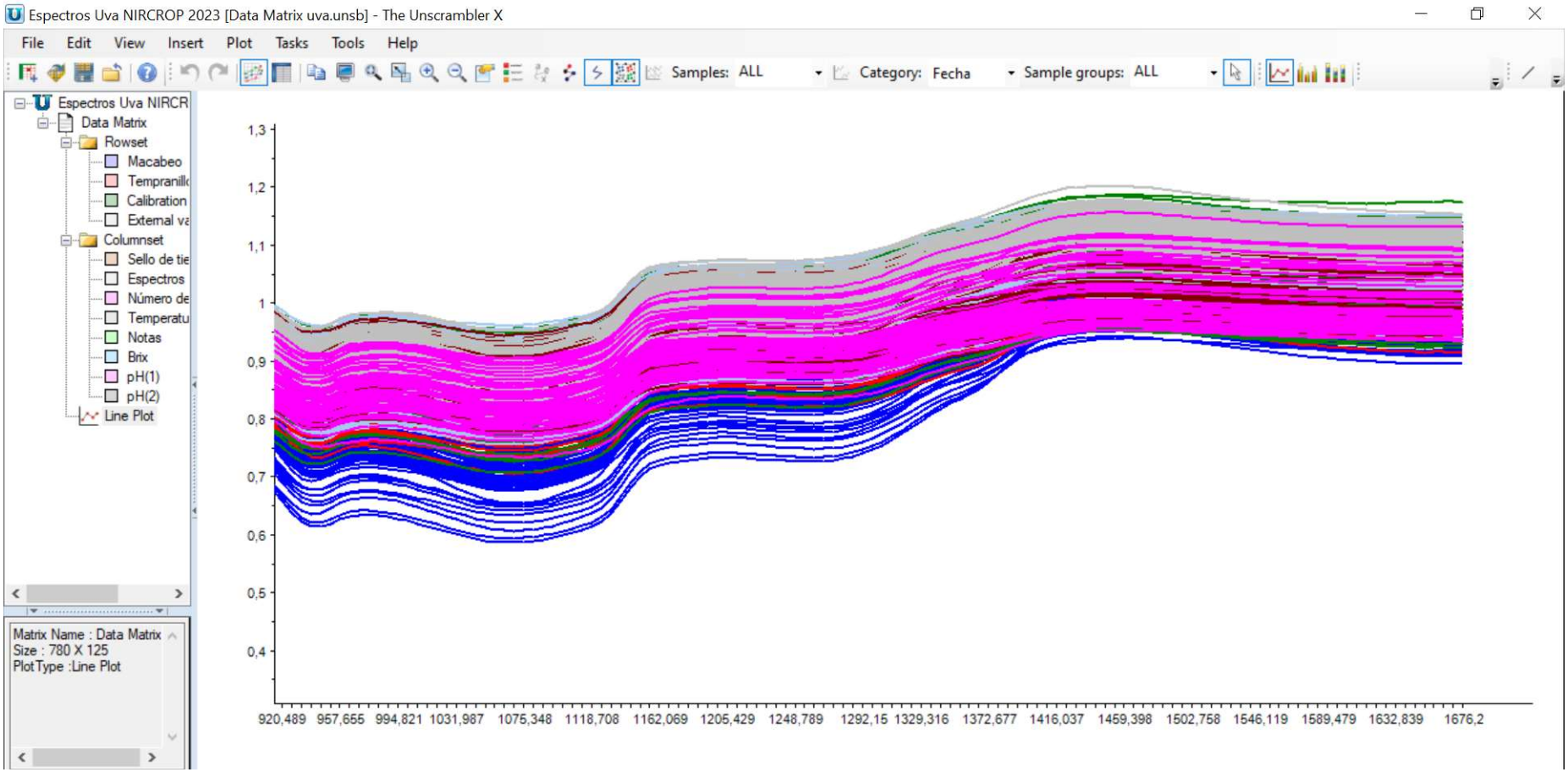
In order to visualize the spectra and detect possible anomalies

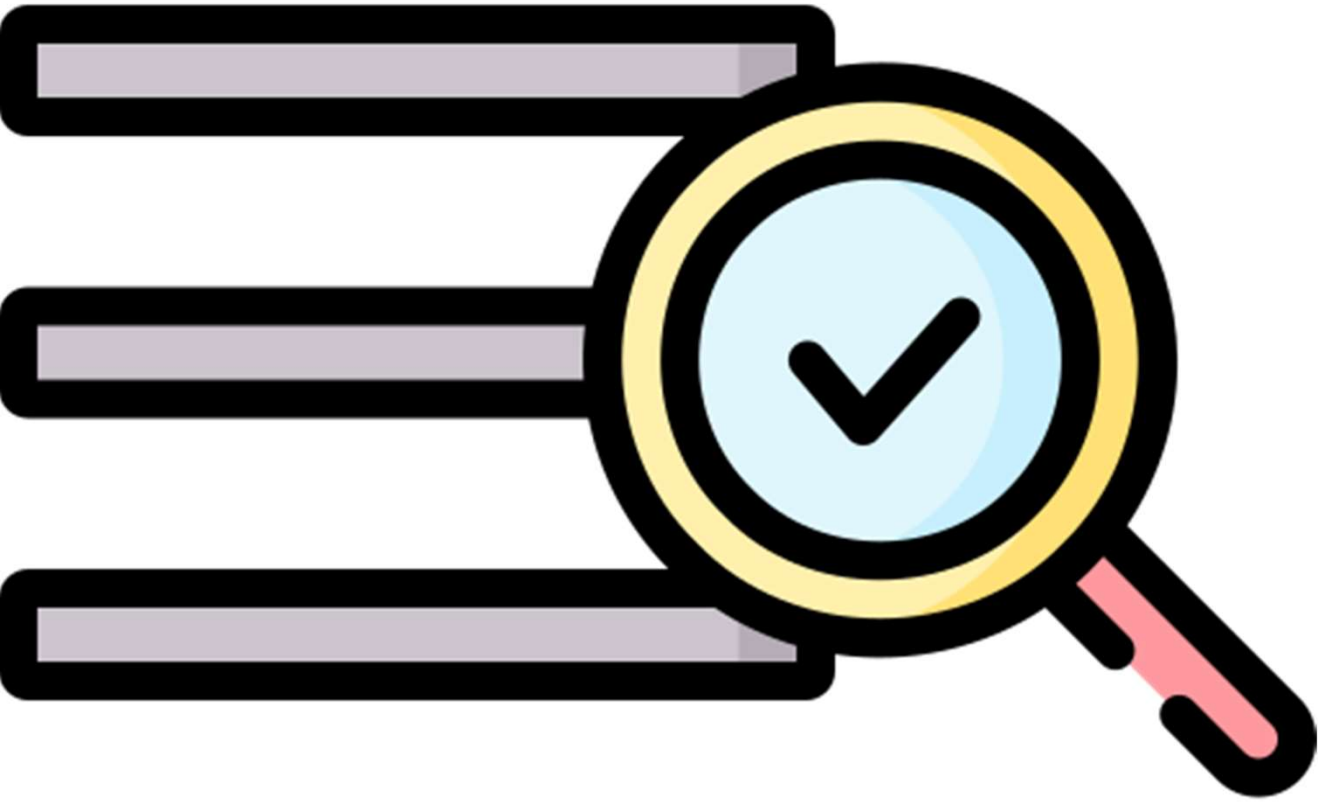
Plot/line





- Plot line


Espectros Uva NIRCROP 2023 [Data Matrix uva.unsb] - The Unscrambler X
File Edit View Insert Plot Tasks Tools Help
Samples: ALL
Category: Fecha
Sample groups: ALL
Espectros Uva NIRCR
Data Matrix
Rowset
Macabeo
Tempranill
Calibration
External va
Columnset
Sello de tie
Espectros
Número de
Temperatu
Notas
Brix
pH(1)
pH(2)
Line Plot
Matrix Name : Data Matrix
Size : 780 X 125
Plot Type :Line Plot
1,3
1,2
1,1
1
0,9
0,8
0,7
0,6
0,5
0,4
920,489 957,655 994,821 1031,987 1075,348 1118,708 1162,069 1205,429 1248,789 1292,15 1329,316 1372,677 1416,037 1459,398 1502,758 1546,119 1589,479 1632,839 1676,2



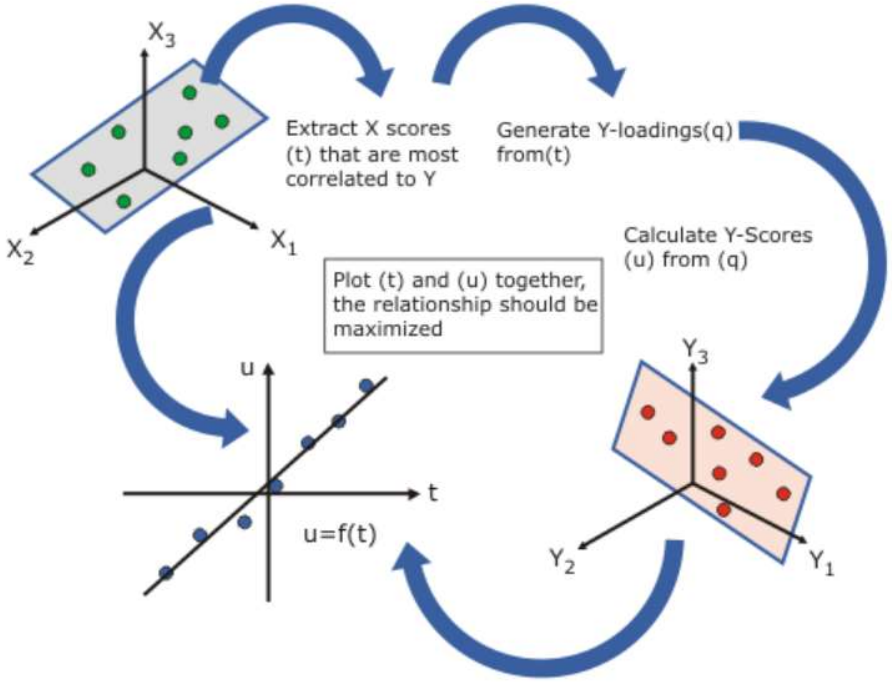
Index

Quantitative models

- **Construction of the model by means of Partial least squares regression (PLSR)**
- Methods available for validation
- Plots
- Statistics for assessing a quantitative model
- Predict
- Outliers

- Construction of the model by means of Partial least squares regression (PLSR)

Partial Least Squares regression models both the X- and Y-matrices simultaneously **to find the latent variables in X that will best predict the latent variables in Y.** These PLSR components are similar to principal components; however, they are referred to as *factors*.
(It finds a linear regression model by projecting the predicted variables and the observable variables to a new space).



- Construction of the model by means of Partial least squares regression (PLSR)

Partial Least Squares regression is a multivariate analysis, i.e. it uses multiple variables - spectral variables - to estimate a given physicochemical parameter.

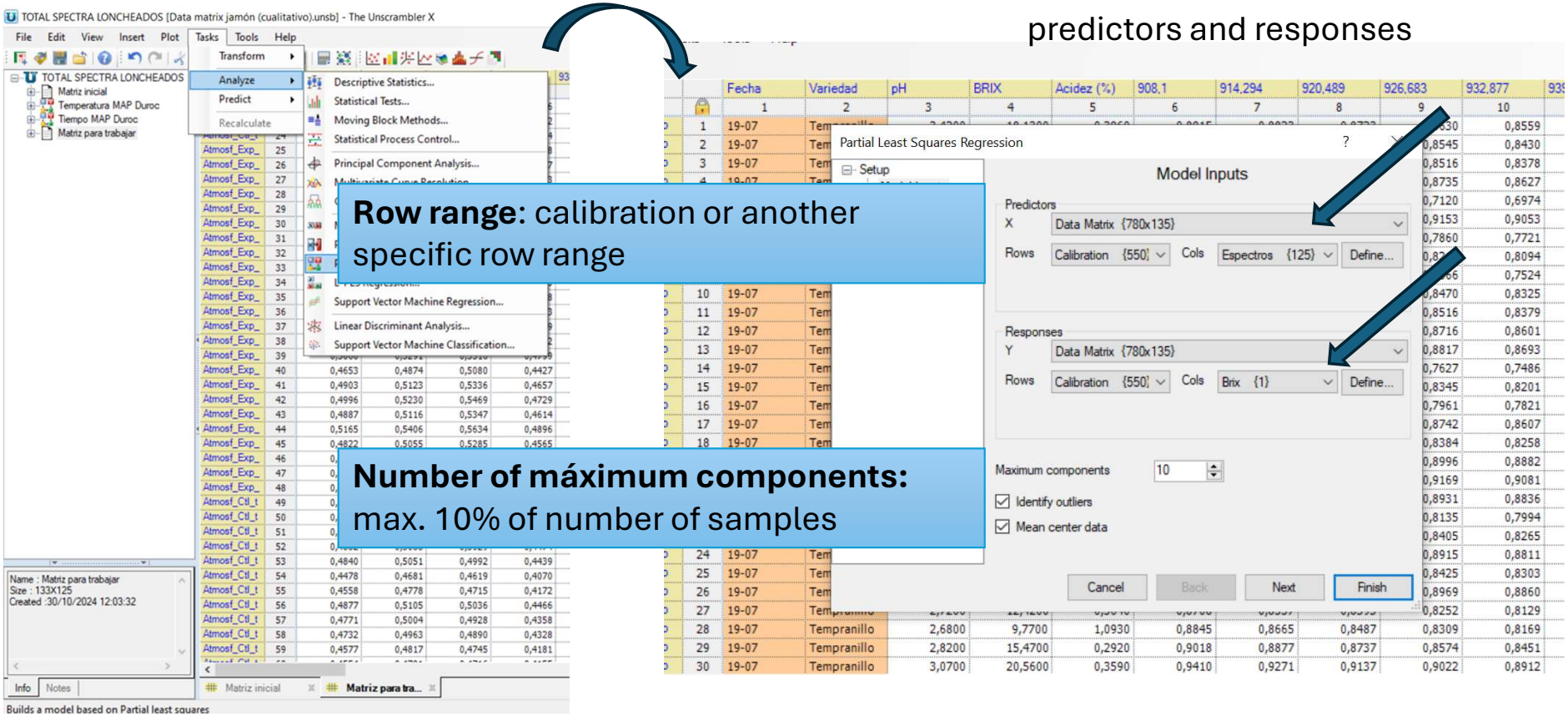
Why cannot a univariate approach be performed in NIRS?

- ❑ **Overlapping absorption bands of different sample constituents**
- ❑ **Large amount of data**
- ❑ **High redundant information (collinearity)**

Causes low selectivity of NIR data

- Construction of the model by means of Partial least squares regression (PLSR)

Task/Analyze/Partial least square regression

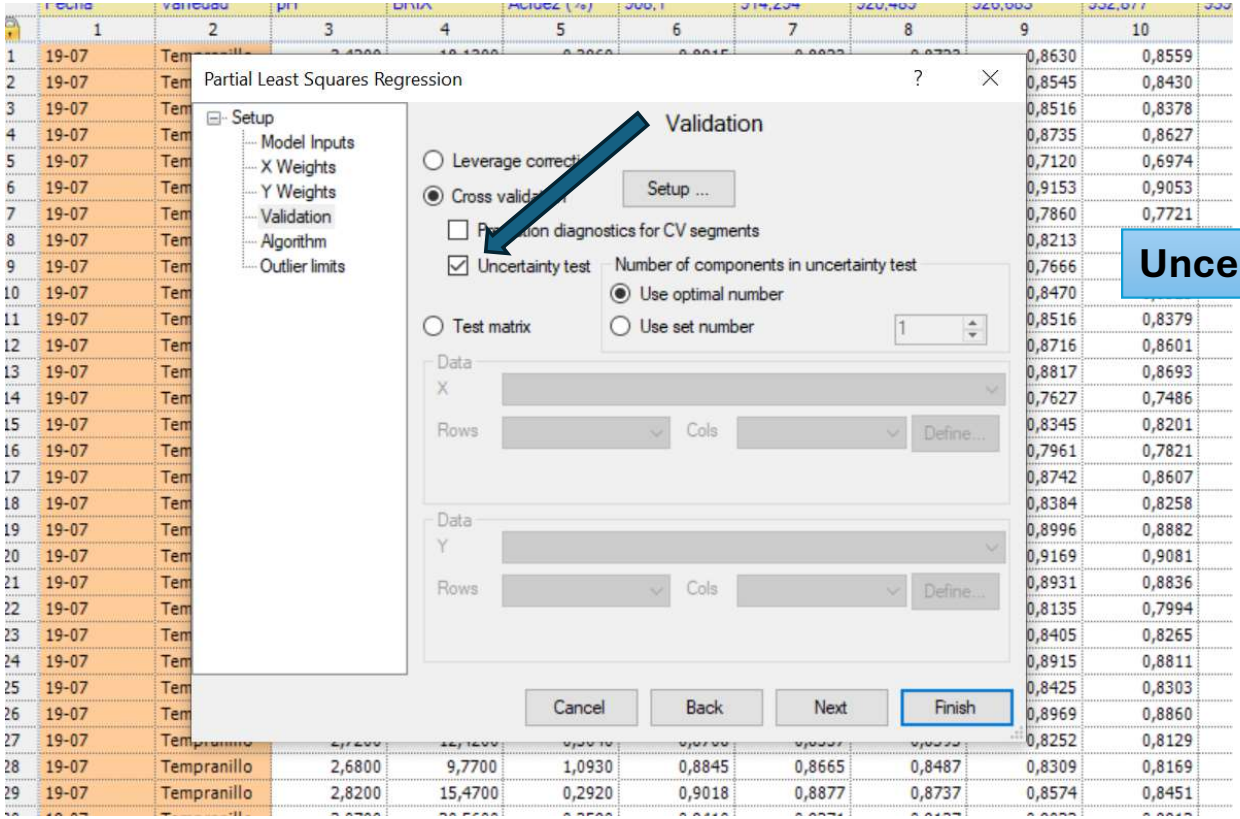


Row range: calibration or another specific row range

Number of máximum components: max. 10% of number of samples

Data matrix: the same for predictors and responses

- Construction of the model by means of Partial least squares regression (PLSR)

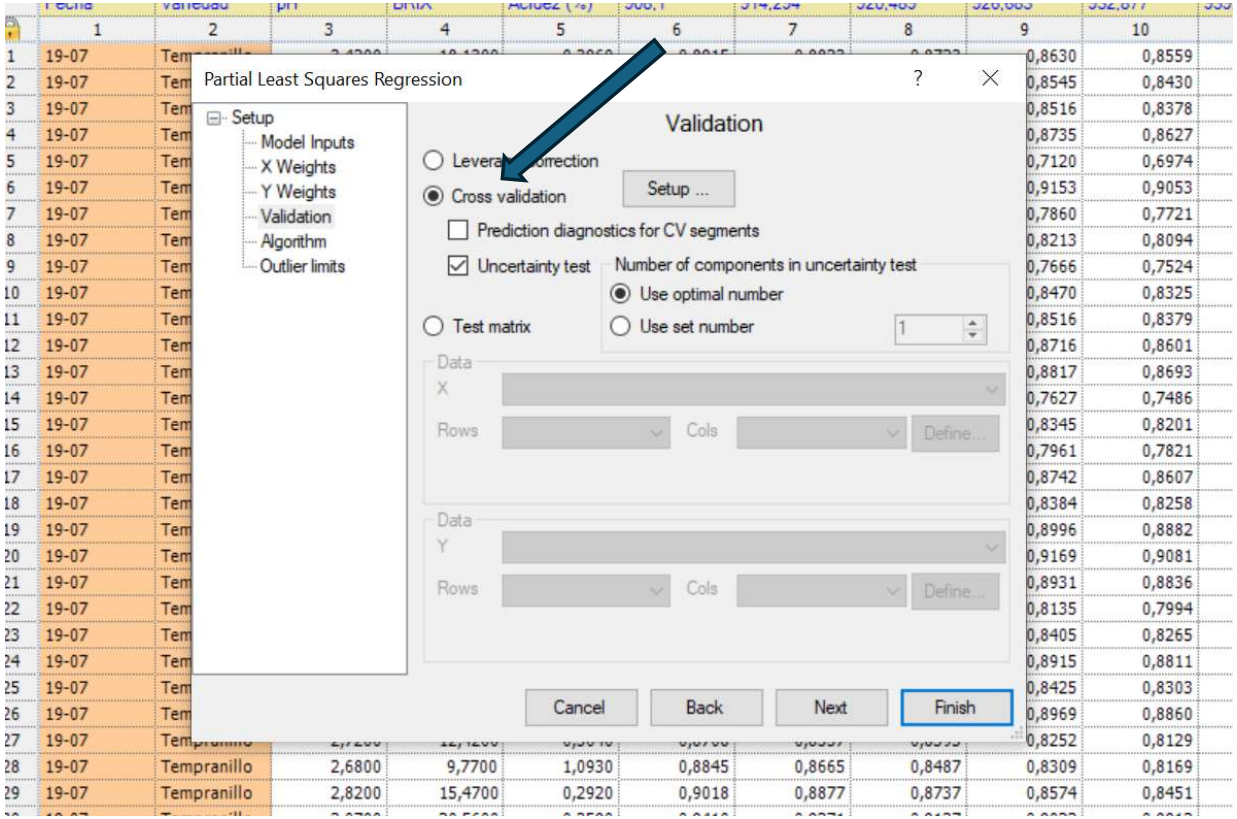


Uncertainty test/Use optimal number

Quantitative models

- Construction of the model by means of Partial least squares regression (PLSR)
- **Methods available for validation**
- Plots
- Statistics for assessing a quantitative model
- Predict
- Outliers

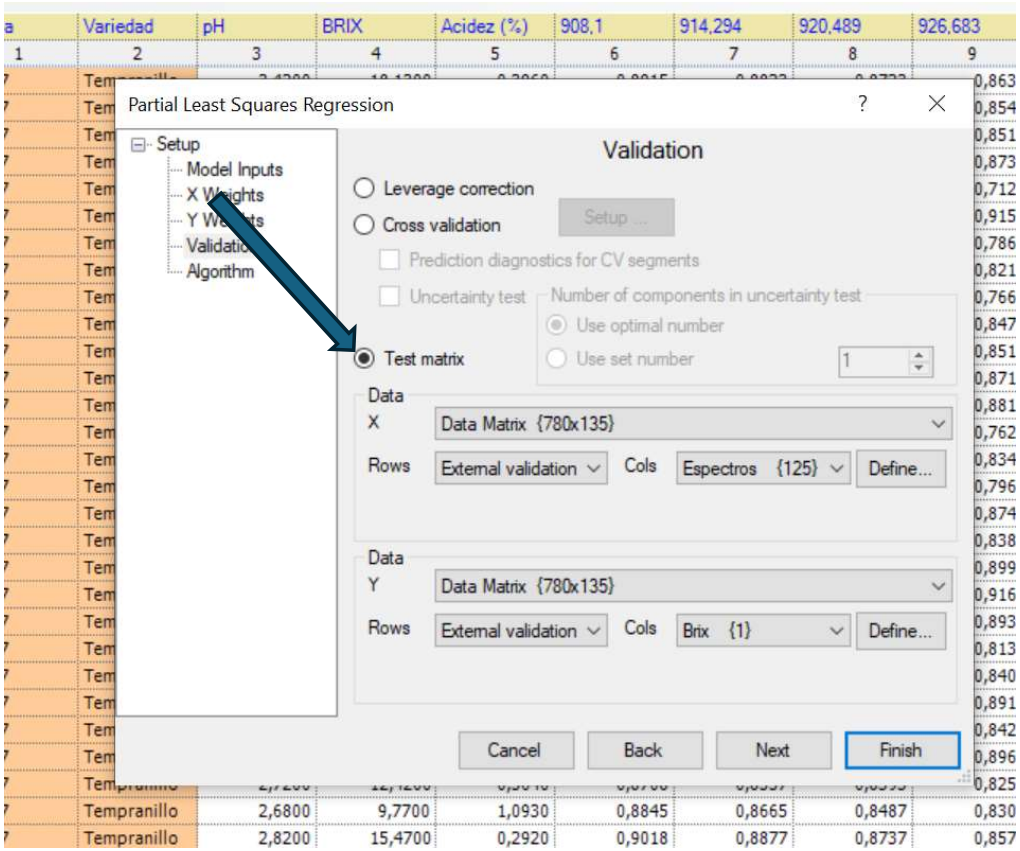
- Methods available for validation



Cross-Validation

This method is used when either **there are not enough samples available to make a separate test** set, or for simulating the effects of different validation test cases, e.g. systematically leaving samples out vs. randomly leaving samples out, etc.

- Methods available for validation



Test matrix

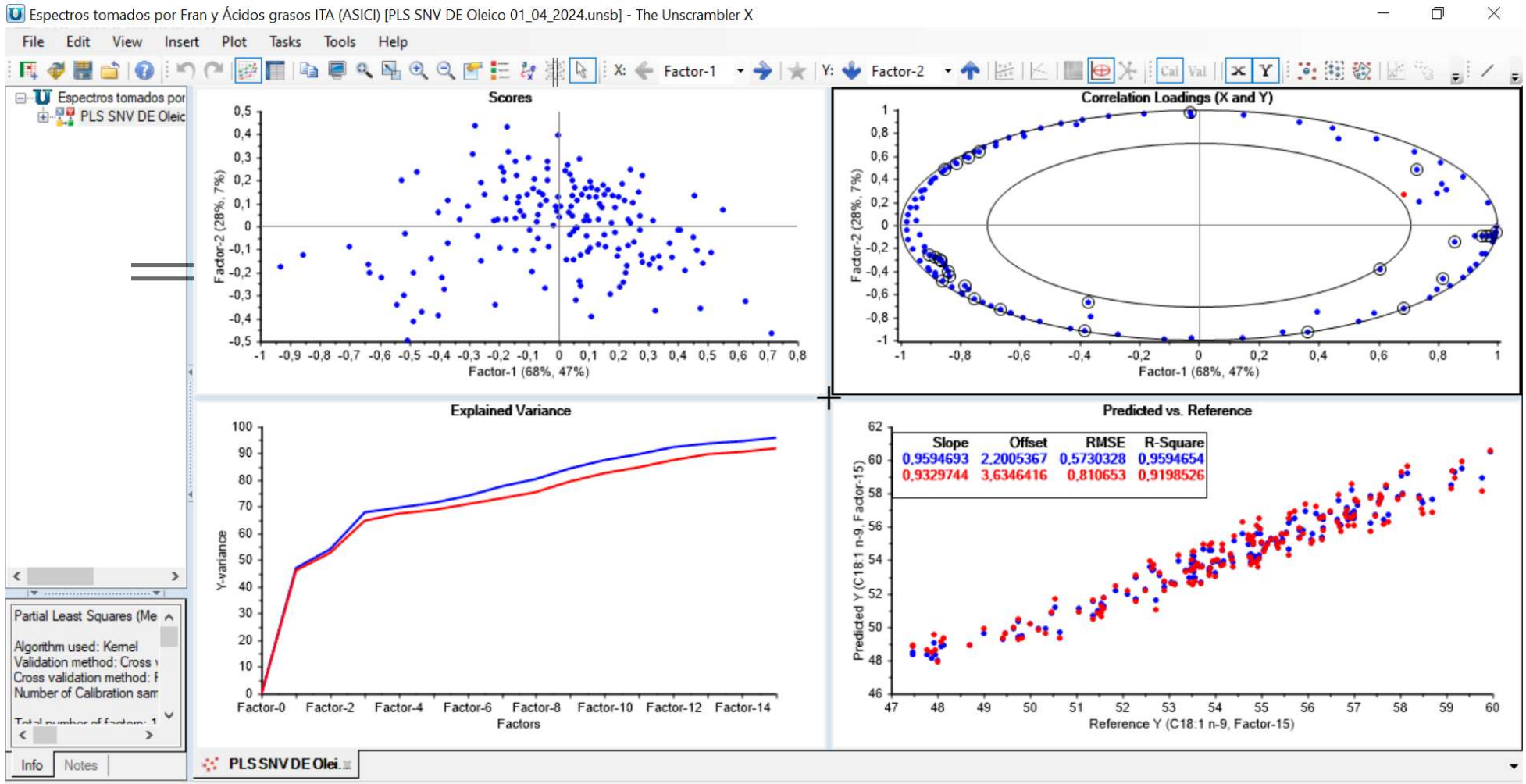
This is also known as *Test Set Validation and* **uses independent samples that have not taken part in the calibration** for validation. Both X- and Y-matrices need to be defined in this case. **This is the preferred method for validation and should be aimed for.**

Quantitative models

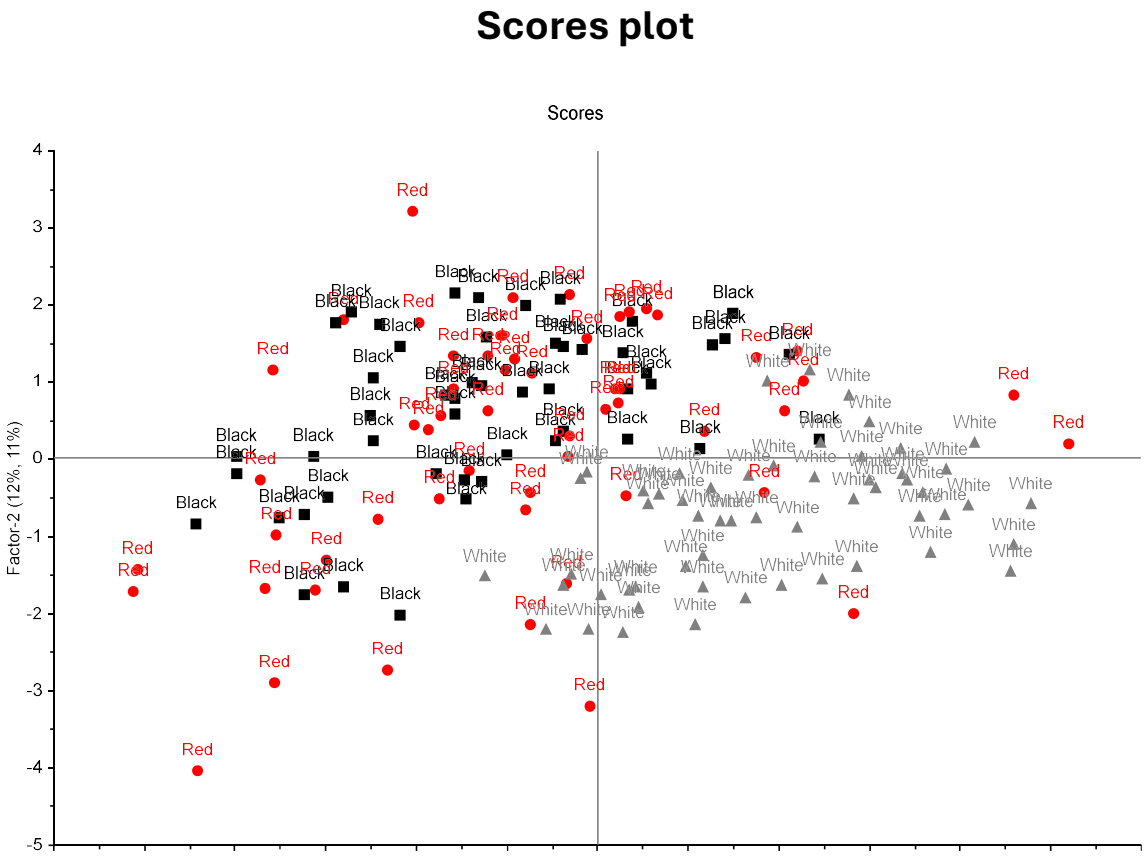
- Construction of the model by means of Partial least squares regression (PLSR)
- Methods available for validation
- **Plots**
- Statistics for assessing a quantitative model
- Predict
- Outliers

- Plots

Regression Overview



- Plots

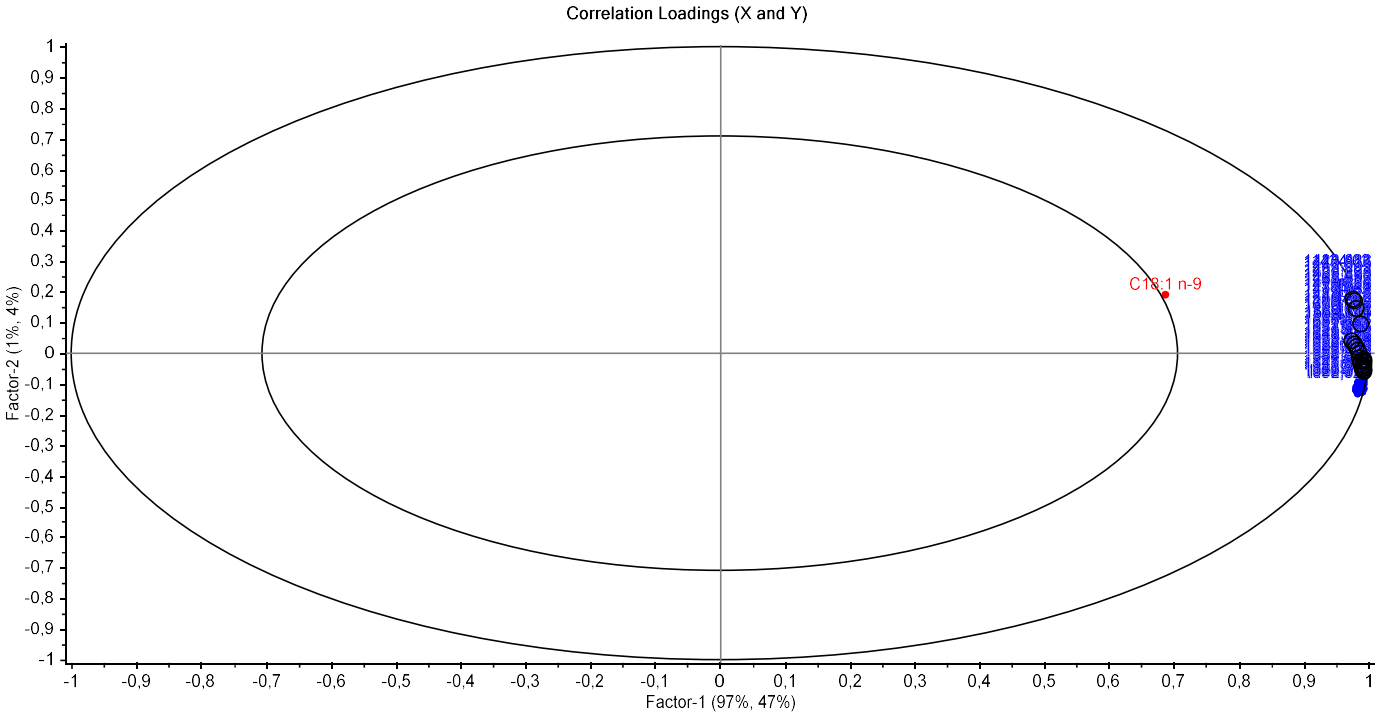


This is a two-dimensional scatter plot (or map) of scores for two specified factors (latent variables or PCs) from PLS regression.

The **plot gives information about patterns** in the samples. The scores plot for (factor 1,factor 2) is especially useful, since these two components summarize more variation in the data than any other pair of components**. The closer the samples are in the scores plot, the more similar they are with respect to the two components concerned**. Conversely, samples far away from each other are different from each other, so the plot can be used to interpret differences and similarities among samples. **It allows to study sample distribution or even detect an outlier**.

- Plots

Correlation loadings

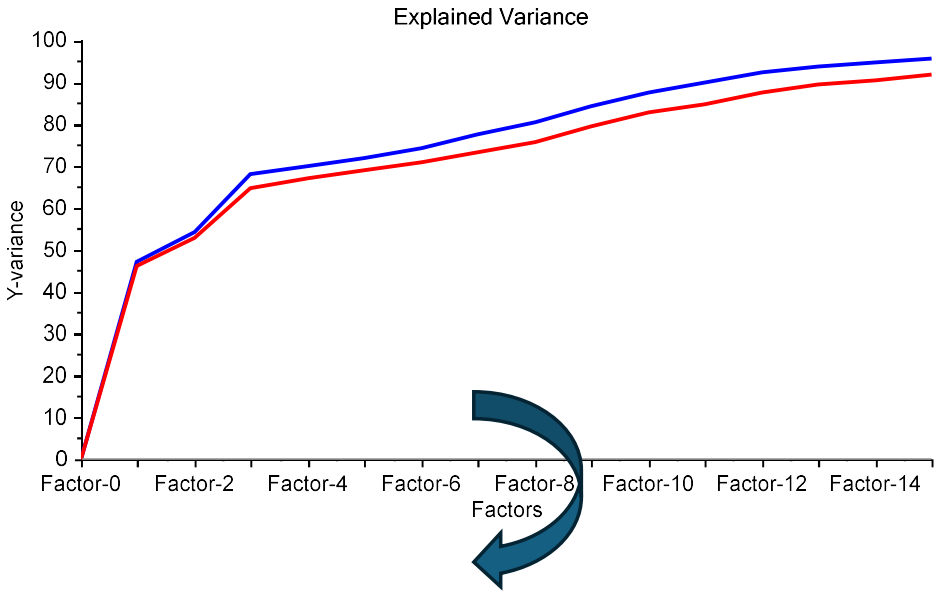


A 2-D scatter plot of X- and Y-loadings for two specified components (factors) from PLS **is a good way to detect important variables and relationships between variables**. The two ellipses help to check how much variance is taken into account. The outer ellipse indicates 100% explained variance. The inner ellipse indicates 50% of explained variance.

- Predictors (X) projected in roughly the same direction from the center as a response, are positively linked to that response.
- Predictors projected in the opposite direction have a negative link.
- **Predictors projected close to the center, are not well represented in that model and cannot be interpreted**.

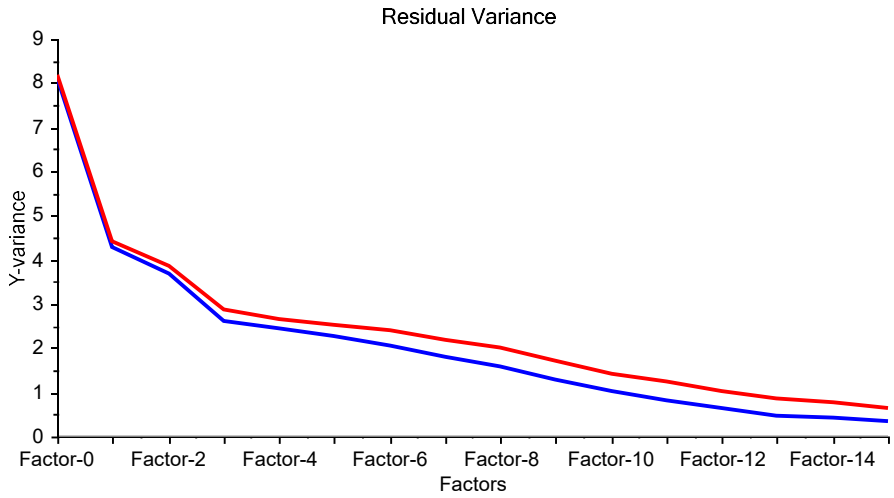
- Plots

Explained/Residual variance variance



This plot illustrates how much of the variation in the response is described by each different component. **It is the percentage of the original variance in the data that is taken into account by the model**

Residual variance

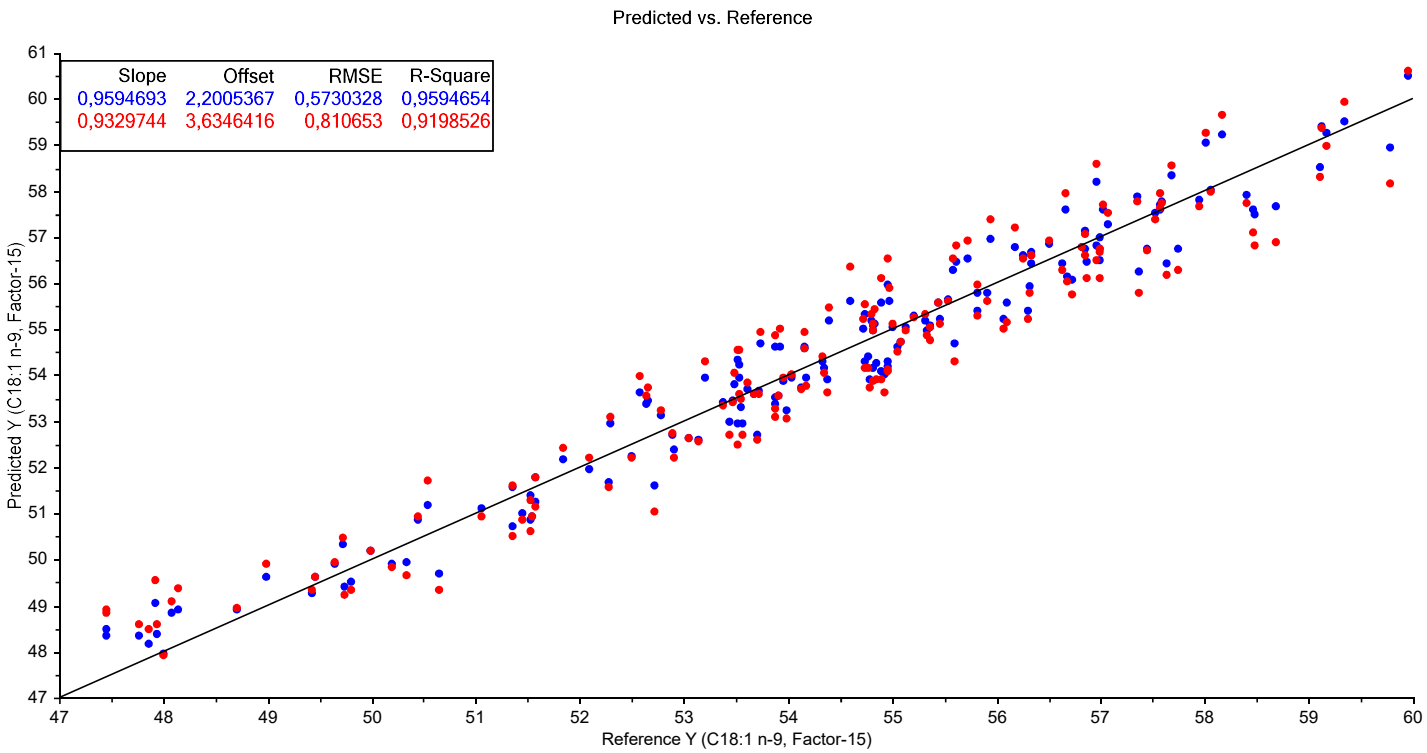


Important!!!

Compare the two variances (calibration and validation): **if they differ significantly, there is good reason to question whether either the calibration data or the test data are truly representative!!**

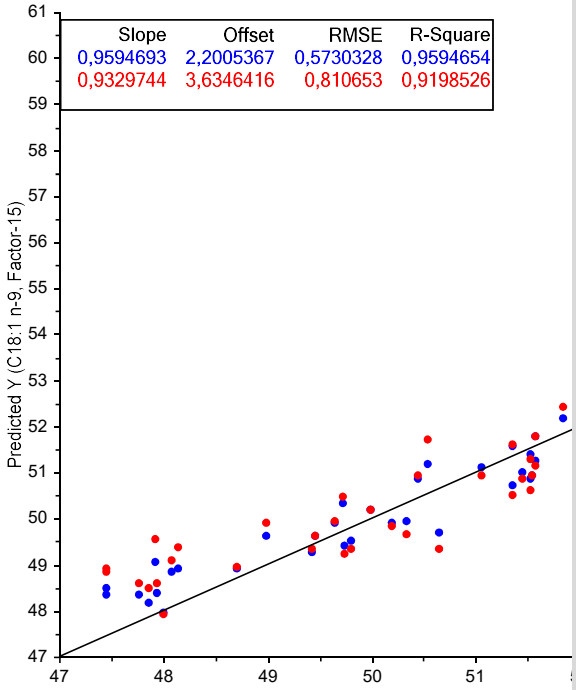
- Plots

Predicted vs. Reference



The predicted Y-value from the model is plotted against the measured Y-value. This is a good way to check the quality of the regression model**. If the model gives a good fit, the plot will show points close to a straight line through the origin and with slope equal to 1**

- Plots



The predicted Y-value from t
check the quality of the regr
close to a straight line throu

<u>To determine the quality of the fit, the following statistics are available</u>,

Slope

The closer the slope is to **1**, the data are better modelled.

Offset

This is the intercept of the line with the Y-axis when the X-axis is set to zero (**Note**: It is not a necessity that this value is zero!)

RMSE

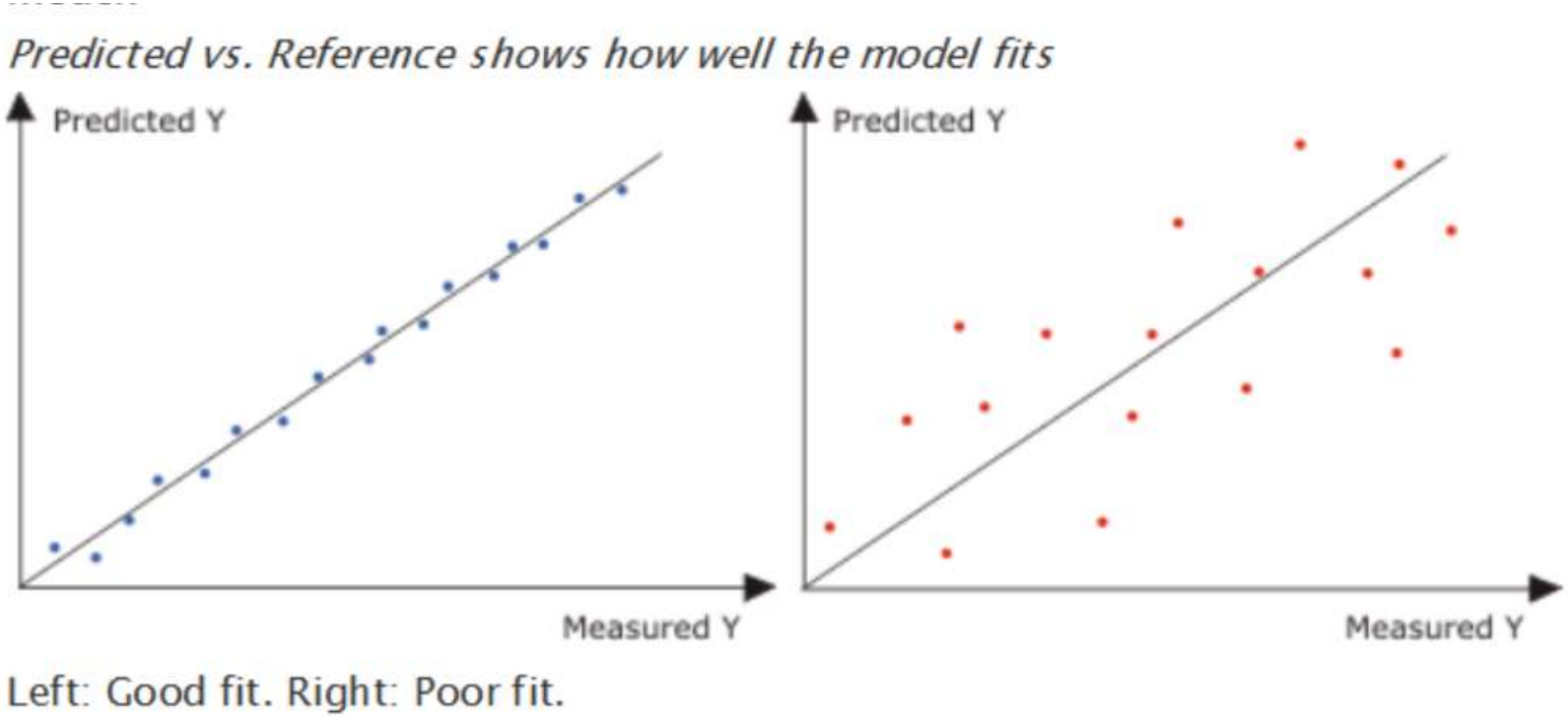
The first one (in blue) is the Calibration error. **This is a measure of the dispersion of the calibration samples about the regression line.** The second one (in red) is the expected Prediction error, depending on the validation method used. Both are expressed in the same unit as the response variable Y.

R-squared

It tells how good a fit can be expected for future predictions for a defined number of components. The first one (in blue) is the calibration, the second one (in red) for the validation set.

- Plots

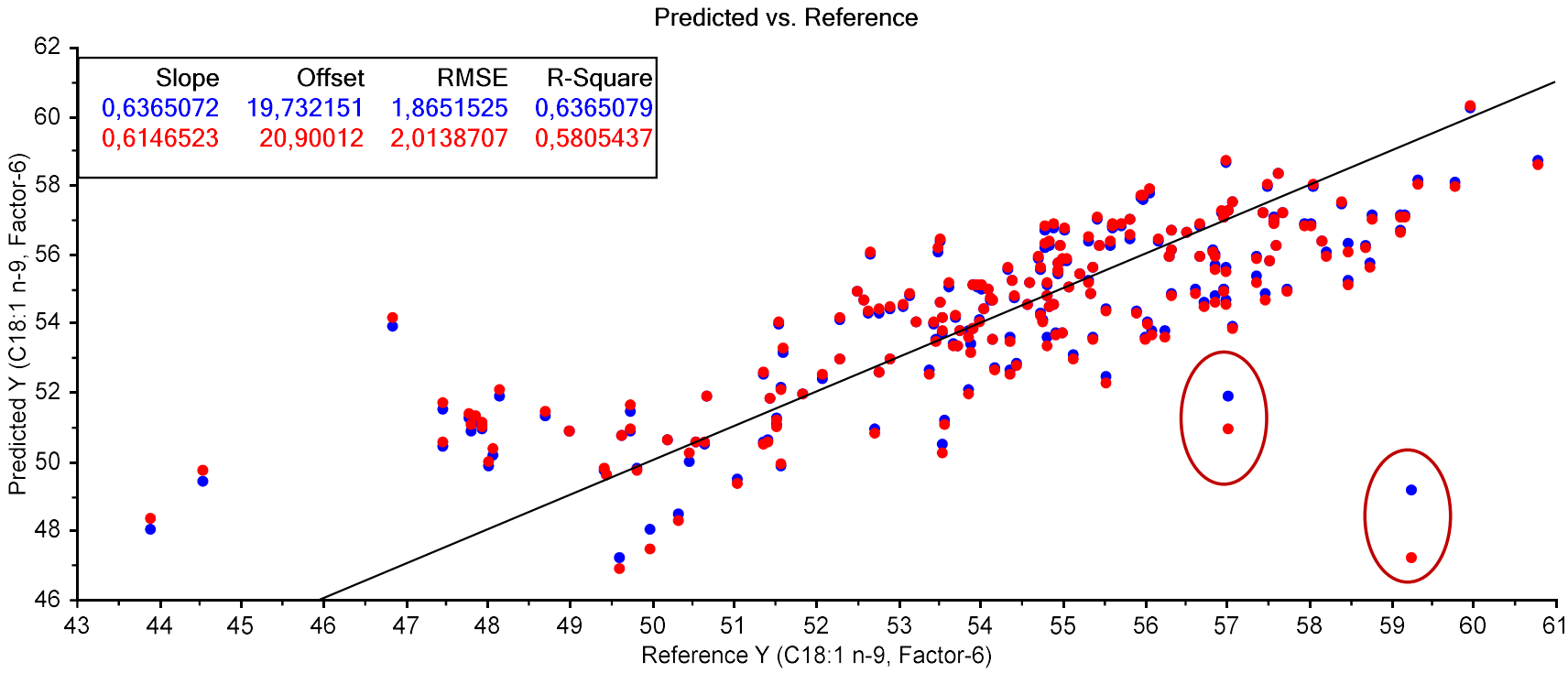
Predicted vs. Reference



- Plots

Predicted vs. Reference

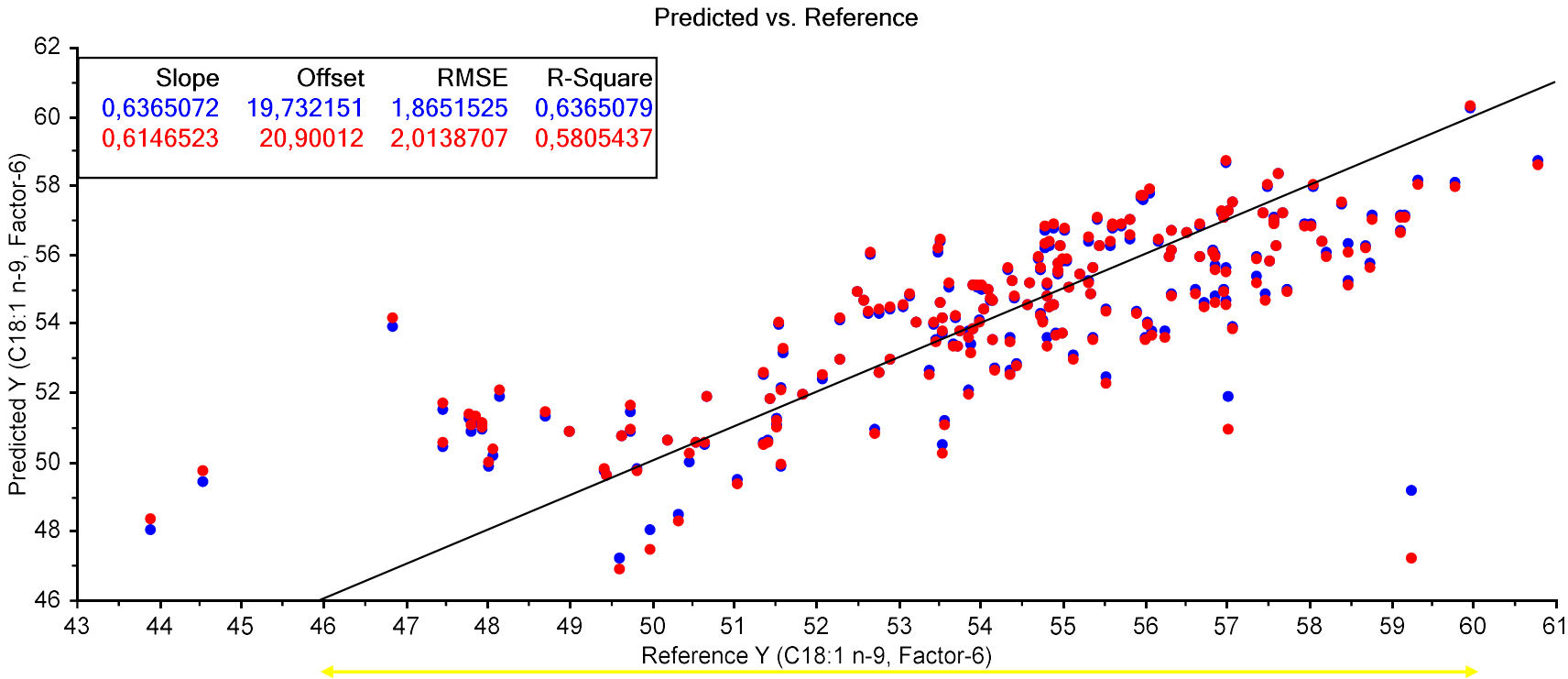
If there are large differences between the calibration and validation results (in cross-validation), the model cannot be trusted.



- Plots

Predicted vs. Reference

To achieve a good (robust, reliable) quantitative model, it is necessary to have a wide range of the variable to be predicted.



Quantitative models

- Construction of the model by means of Partial least squares regression (PLSR)
- Methods available for validation
- Plots
- **Statistics for assessing a quantitative model**
- Predict
- Outliers

- Statistics for assessing a quantitative model

Coeficiente de determinación en validación cruzada (1-VR)

- ❑ 1-VR ≥ 0,90. Excellent calibration
- ❑ 0,89 > 1-VR > 0,70 Good calibration
- ❑ 0,69 > 1-VR > 0,50 The model could be used to discriminate between low, medium and high values.

Error en validación cruzada

Should be as little as possible

Residual Prediction Deviation (RPD)

RPD = SD/RMSE

Routine analisis RPD ≥ 3

Range Error Ratio (RER)

RER = Range/RMSE

- ❑ 4 > RER > 8 The model could be used to discriminate between low, medium and high values
- ❑ 8 > RER > 12 It is possible to predict quantitatively
- ❑ RER > 12 Excellent prediction

Williams, P. (2003). Near Infrared Technology-Getting the best Out of Light. PDK Grain. Nanaimo, Canada

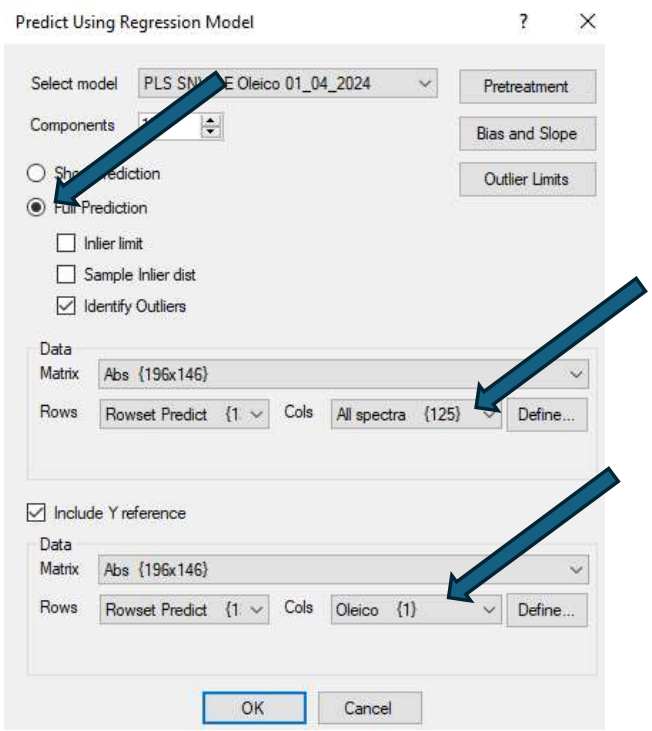
Quantitative models

- Construction of the model by means of Partial least squares regression (PLSR)
- Methods available for validation
- Plots
- Statistics for assessing a quantitative model
- **Predict**
- Outliers

- Predict

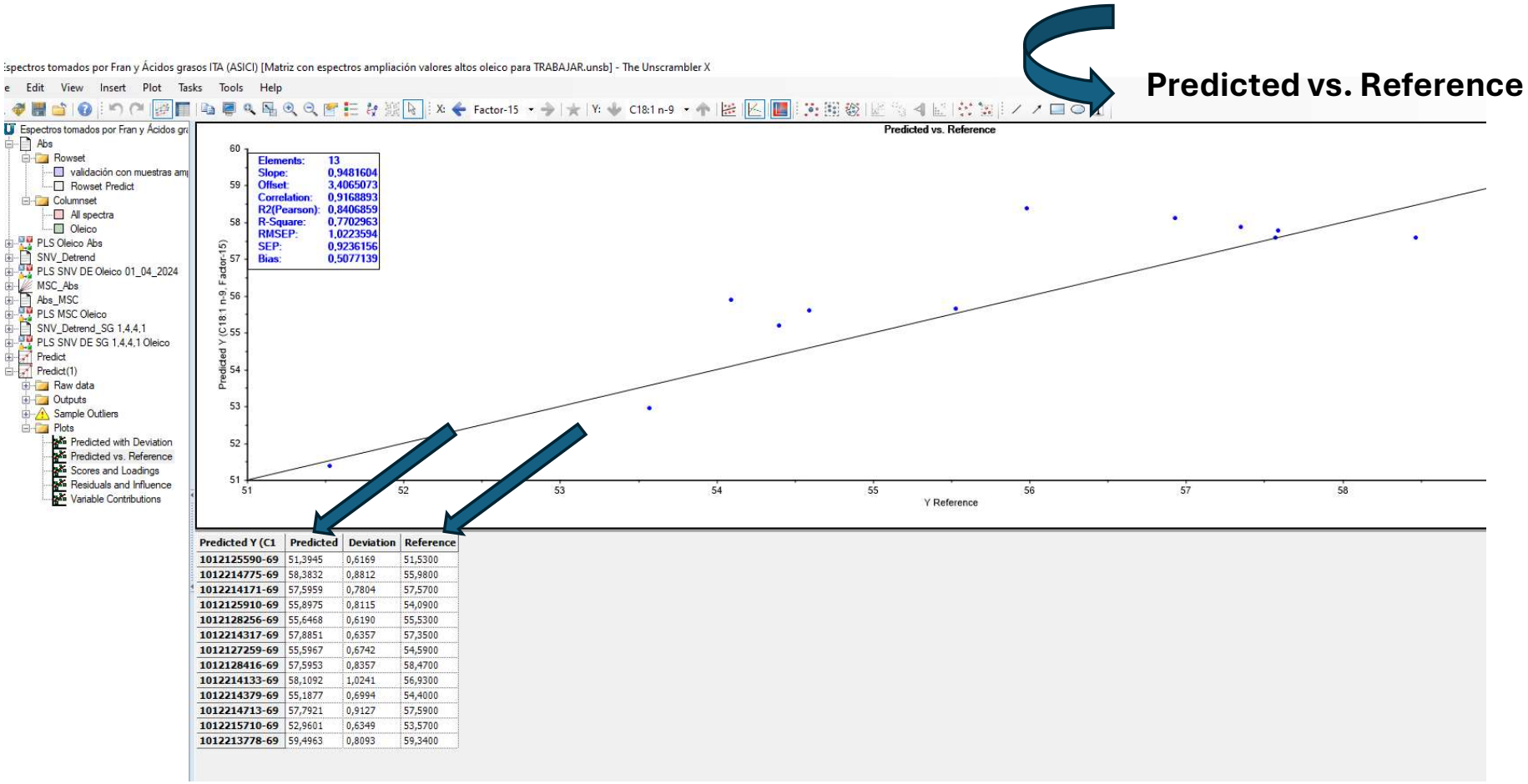
Use an existing regression model to predict response values for new samples.

Tasks/Predict/Regression



- **Full Prediction** uses a projection on the latent space in the calculation. (It will provide comprehensive results such as plots and additional matrices for increased data interpretation and outlier diagnostics).
- **Short Prediction** uses only the extracted Regression (Beta) coefficients. There are no plots associated with this type of prediction.

- Predict



Predicted Y (C1	Predicted	Deviation	Reference
1012125590-69	51,3945	0,6169	51,5300
1012214775-69	58,3832	0,8812	55,9800
1012214171-69	57,5959	0,7804	57,5700
1012125910-69	55,8975	0,8115	54,0900
1012128256-69	55,6468	0,6190	55,5300
1012214317-69	57,8851	0,6357	57,3500
1012127259-69	55,5967	0,6742	54,5900
1012128416-69	57,5953	0,8357	58,4700
1012214133-69	58,1092	1,0241	56,9300
1012214379-69	55,1877	0,6994	54,4000
1012214713-69	57,7921	0,9127	57,5900
1012215710-69	52,9601	0,6349	53,5700
1012213778-69	59,4963	0,8093	59,3400

Quantitative models

- Construction of the model by means of Partial least squares regression (PLSR)
- Methods available for validation
- Plots
- Predict
- **Outliers**

- Outliers

An **outlier is an object which deviates from the other objects in a model and may not belong to the same population as the majority and therefore can disturb the model**.

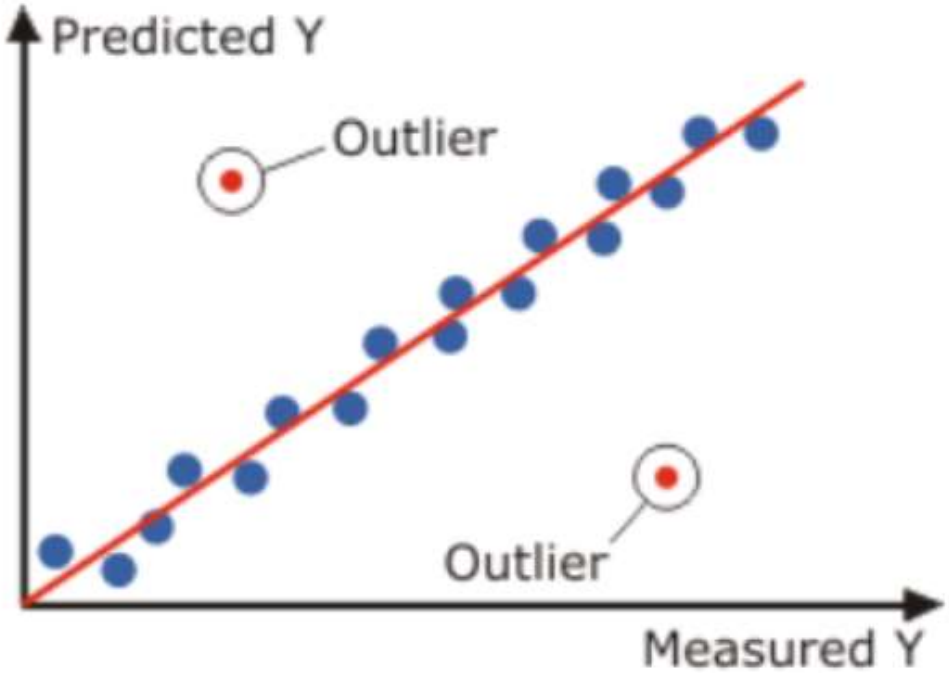
The cause of outliers could be one or more of the following:

- ❖ Measurement error
- ❖ Wrong labeling
- ❖ Deviating products / processes
- ❖ Noise
- ❖ **Extreme / interesting sample**

For projection methods like PLSR, outliers can be detected using scores plots, residuals, leverages and influence plots

- Outliers. **How to detect them?**

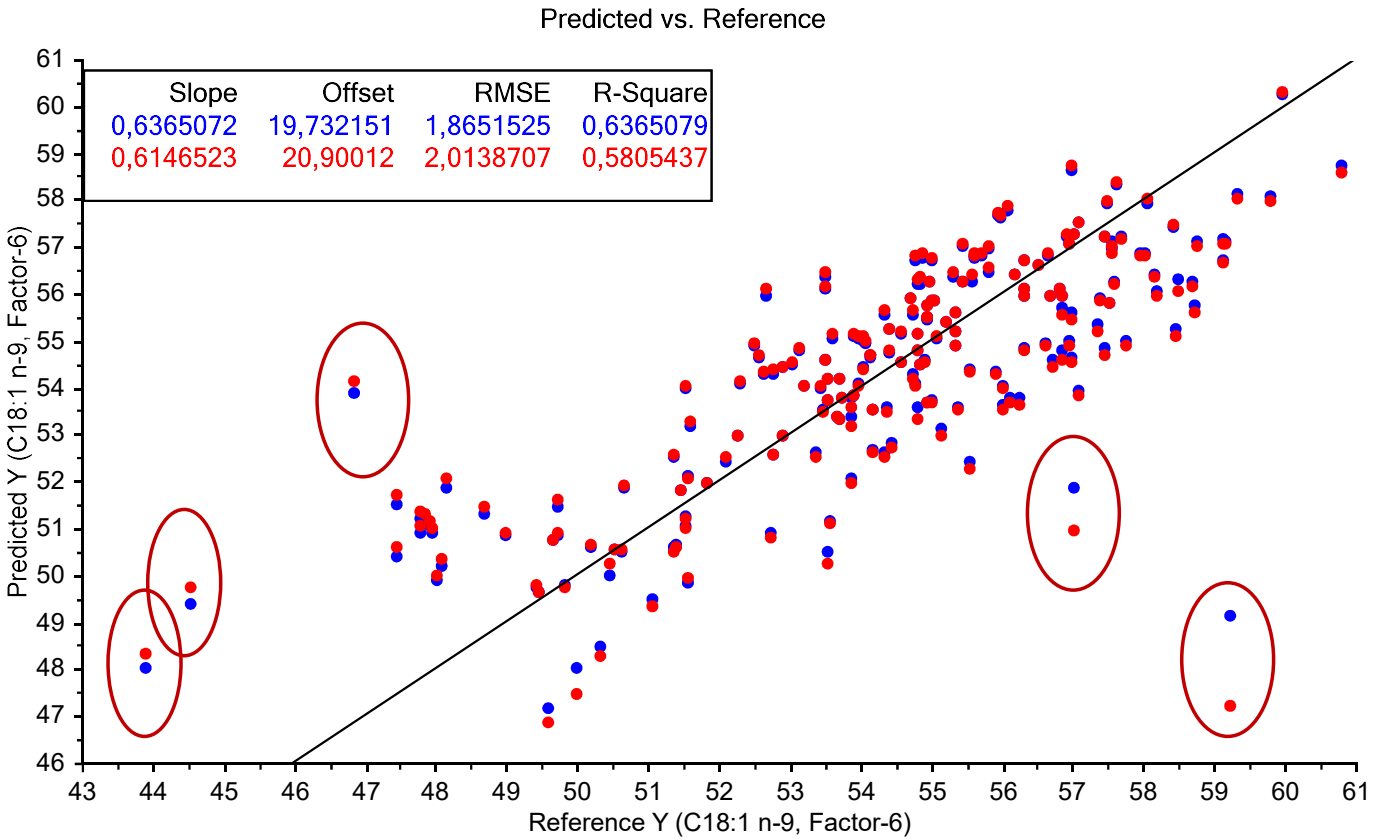
Predicted vs. Reference



One may also see cases where the majority of the samples lie close to the line while a few of them are further away. This may indicate **good fit of the model to the majority of the data, but with a few outliers present**

- Outliers. **How to detect them?**

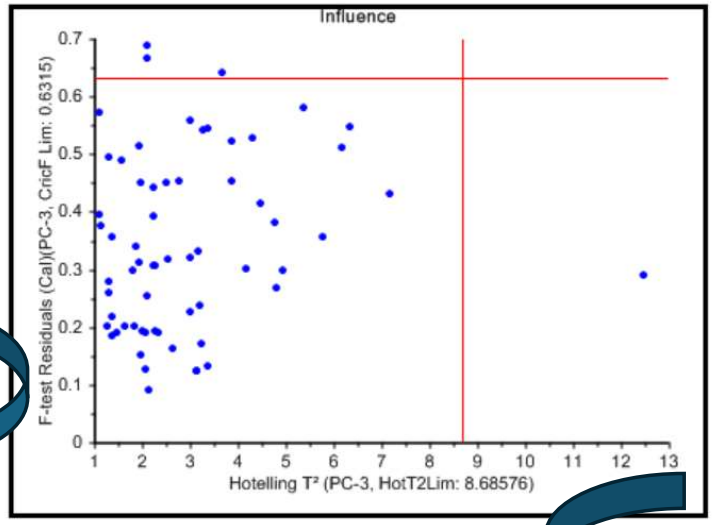
Predicted vs. Reference



- Outliers. **How to detect them?**

Influence plot with Hotelling's T^2 statistic → Representa dos tipos de outliers

Influence plot with Hotelling's T² on the abscissa and F-residuals on the ordinate



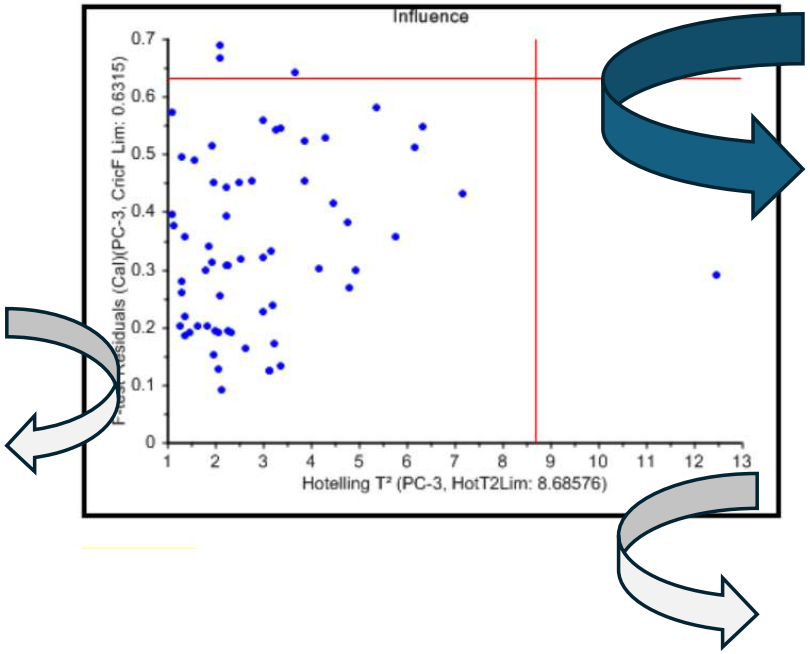
El eje de ordenadas **describe la distancia de la muestra al modelo.**
Las muestras con una varianza residual elevada están mal descritas por el modelo.

Las muestras con un valor elevado en el eje de abscisa (a la derecha) **están bien descritas por el modelo.** Sin embargo, las puntuaciones de las muestras pueden tener valores muy altos o bajos para algunos componentes en comparación con el resto de las muestras. Estas muestras son peligrosas en la fase de calibración porque influyen en el modelo. Una muestra suficientemente extrema puede abarcar por sí sola un componente entero, en cuyo caso el modelo dejará de ser fiable. **Si la varianza descrita por la muestra es importante pero única, hay que intentar obtener más muestras del mismo tipo para estabilizar el modelo**. **De lo contrario, la muestra debe descartarse como valor atípico.**

- Outliers. **How to detect them?**

Influence plot with Hotelling's T^2 statistic → Representa dos tipos de outliers

Influence plot with Hotelling's T² on the abscissa and F-residuals on the ordinate



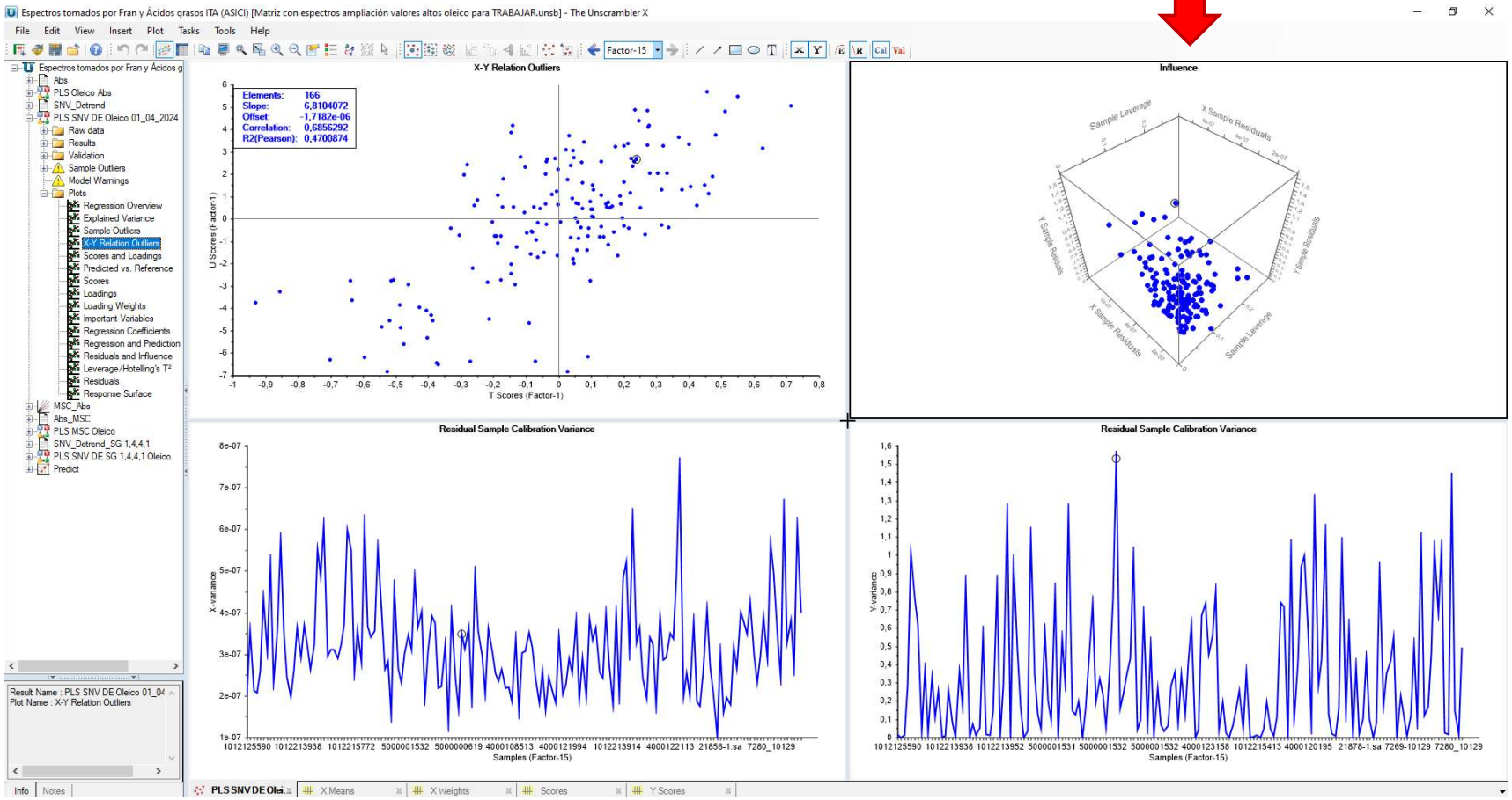
Obsérvese que una muestra con una varianza residual y un valor de Hotelling elevados es el valor atípico más peligroso. No sólo está mal descrita por el modelo, sino que además es influyente.

El eje de ordenadas **describe la distancia de la muestra al modelo.**
Las muestras con una varianza residual elevada están mal descritas por el modelo.

componentes en comparación con el resto de las muestras. Estas muestras son peligrosas en la fase de calibración porque influyen en el modelo. Una muestra suficientemente extrema puede abarcar por sí sola un componente entero, en cuyo caso el modelo dejará de ser fiable. Si la varianza descrita por la muestra es importante pero única, hay que intentar obtener más muestras del mismo tipo para estabilizar el modelo. De lo contrario, la muestra debe descartarse como valor atípico.

- Outliers. **How to detect them?**

Influence plot

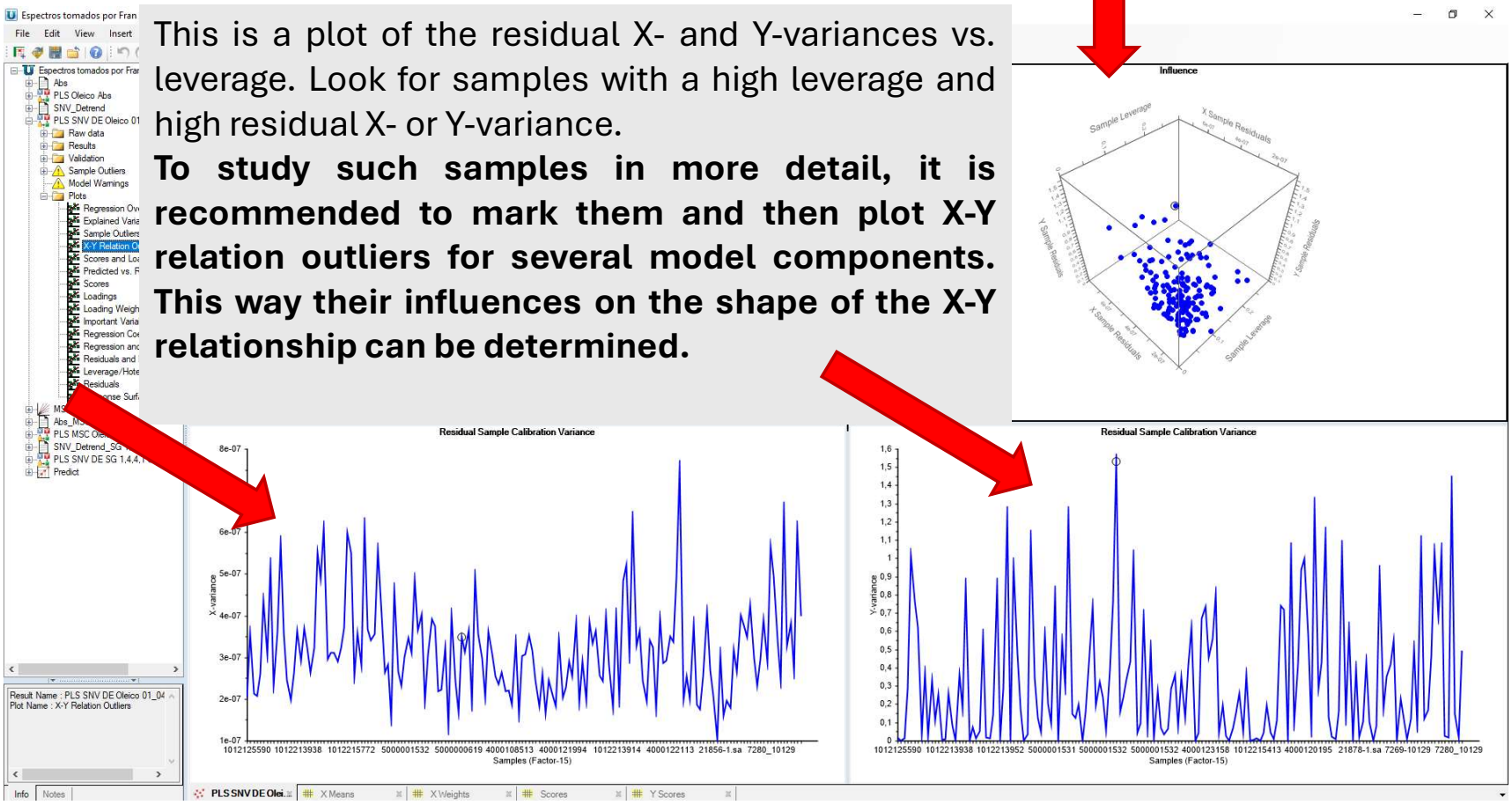


- Outliers. **How to detect them?**

Influence plot

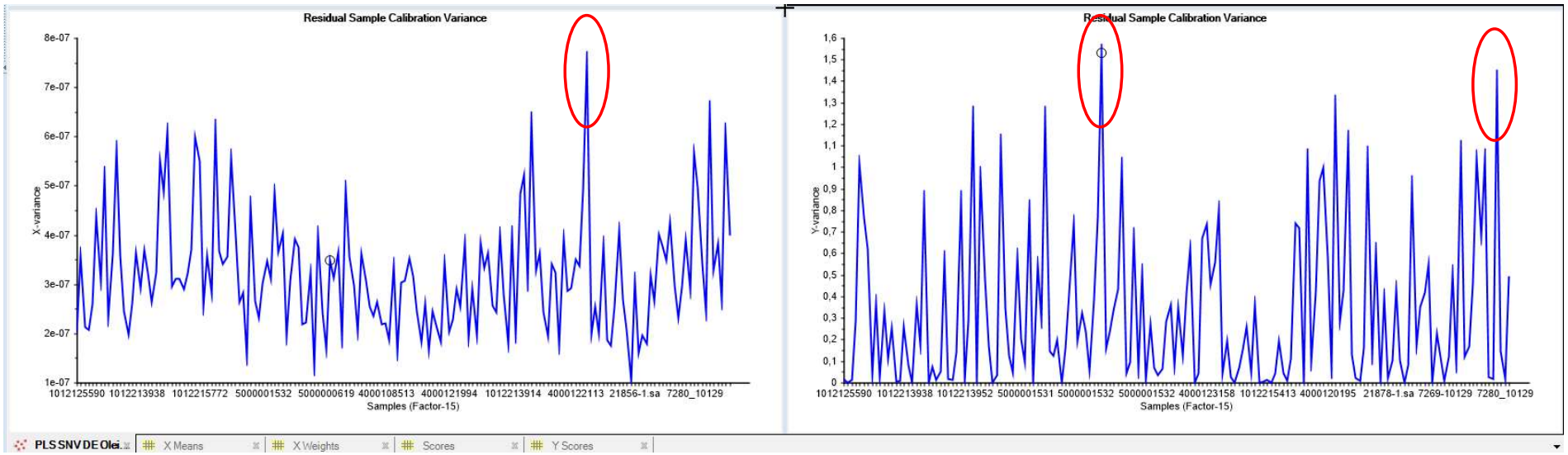
This is a plot of the residual X- and Y-variances vs. leverage. Look for samples with a high leverage and high residual X- or Y-variance.

To study such samples in more detail, it is recommended to mark them and then plot X-Y relation outliers for several model components. This way their influences on the shape of the X-Y relationship can be determined.



- Outliers. **How to detect them?**

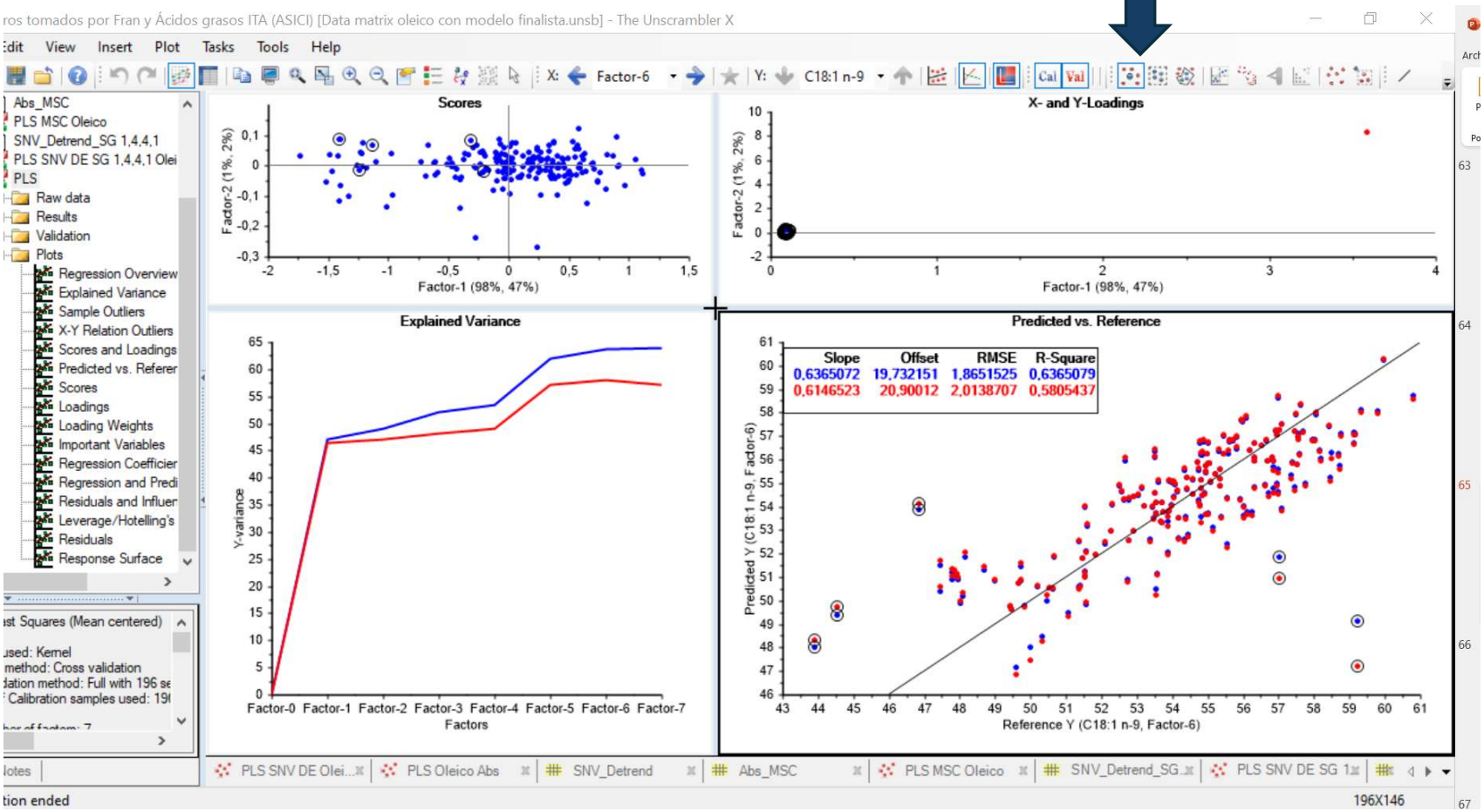
Residual sample variance



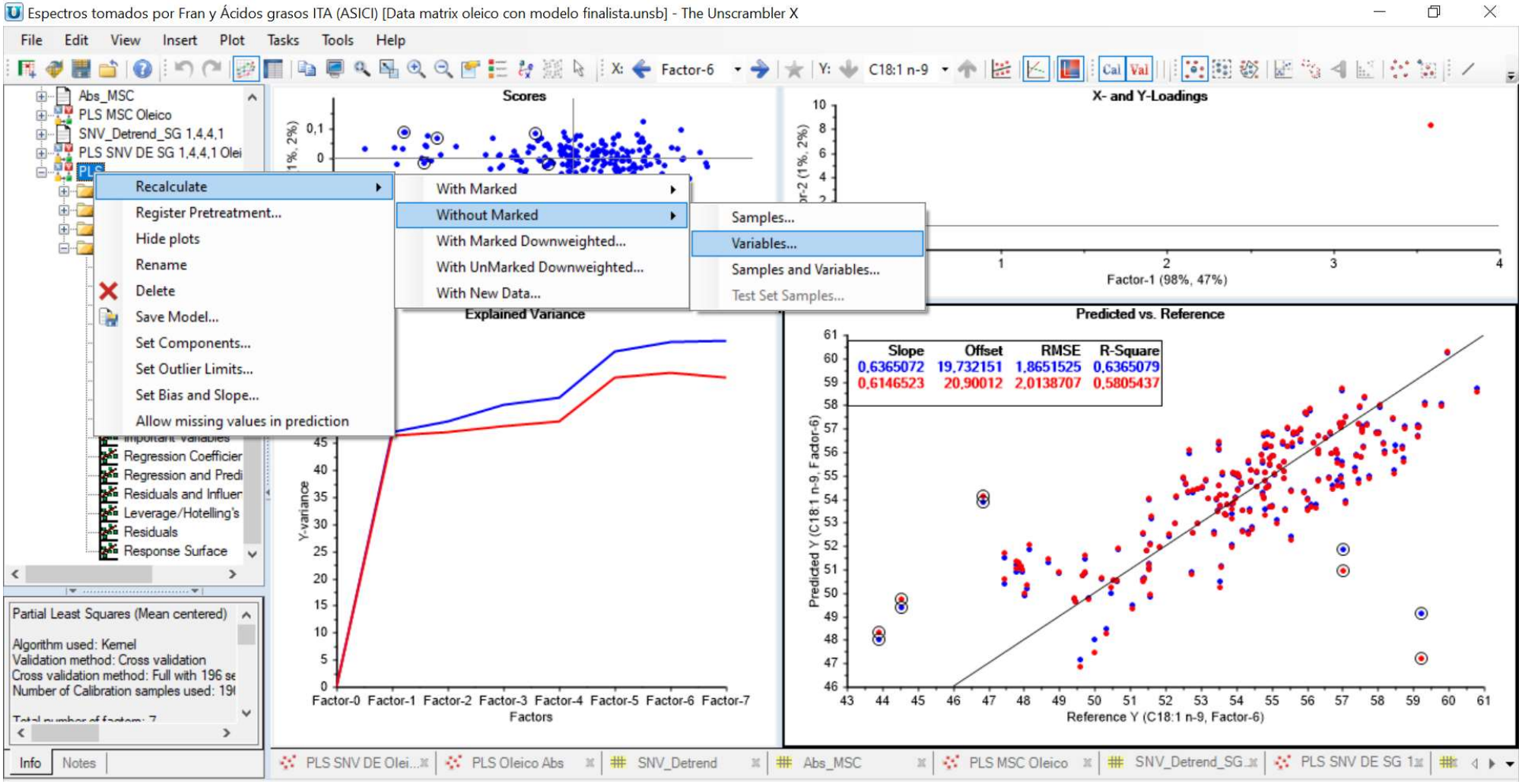
High residuals indicate an outlier. Incorporating more components can sometimes model outliers; avoid doing so since it will reduce the prediction ability of the model.

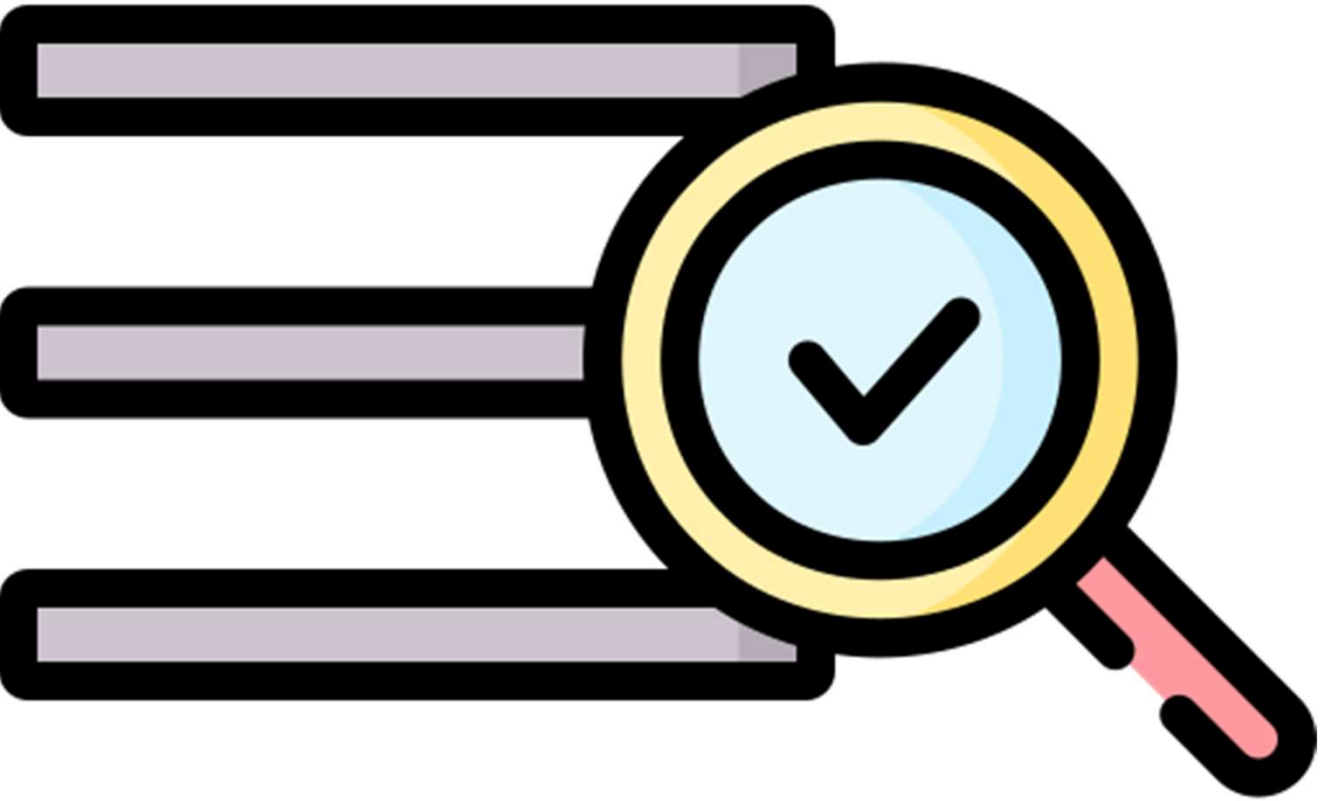
- Outliers. Recalculate the model without outliers

Marc one by one



- Outliers. Recalculate the model without outliers





Index

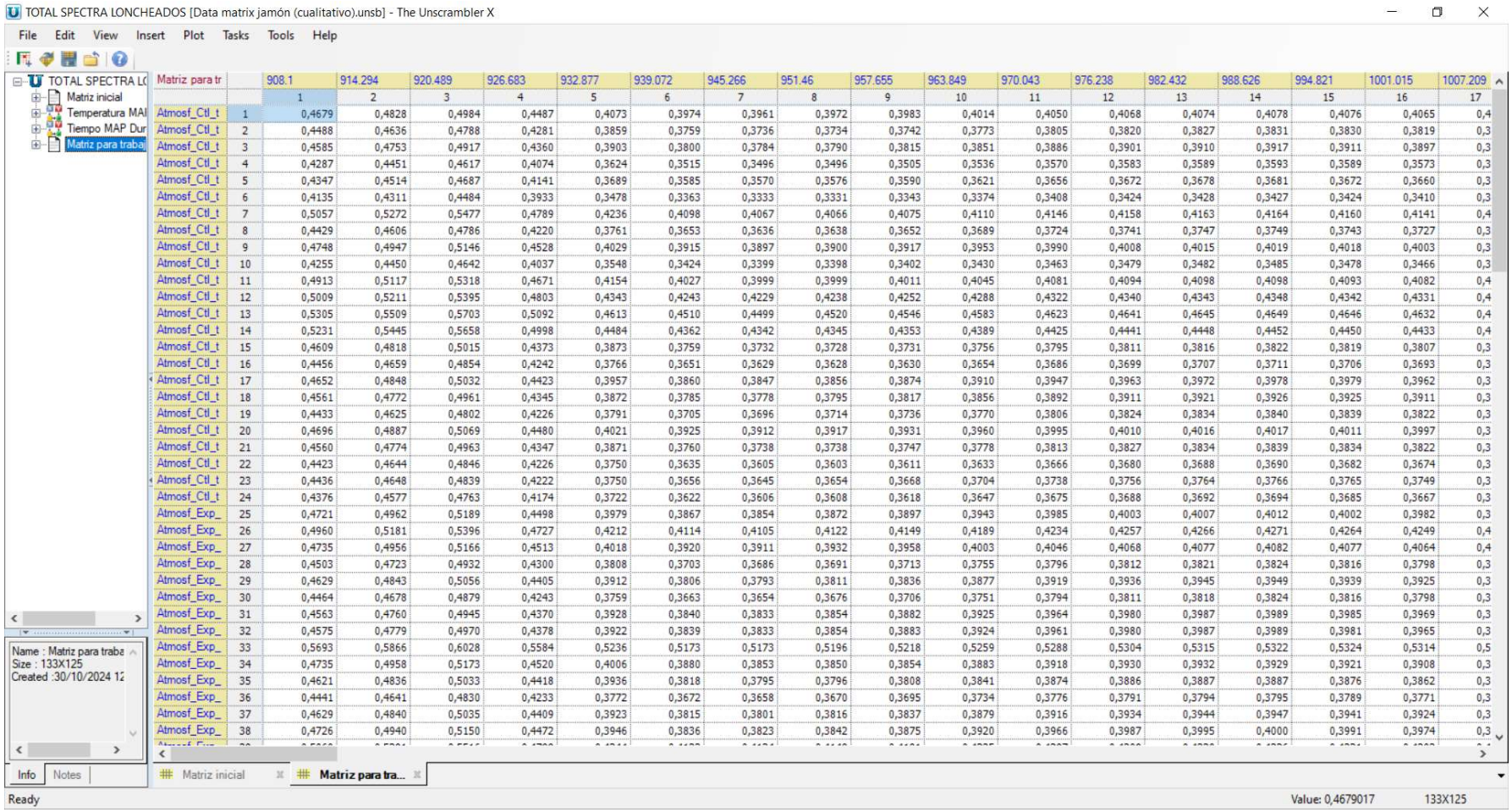
Qualitative models

- Data import
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)
- Development of classificatory models
- Predict

Qualitative models

- **Data import**
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)
- Development of classificatory models
- Predict

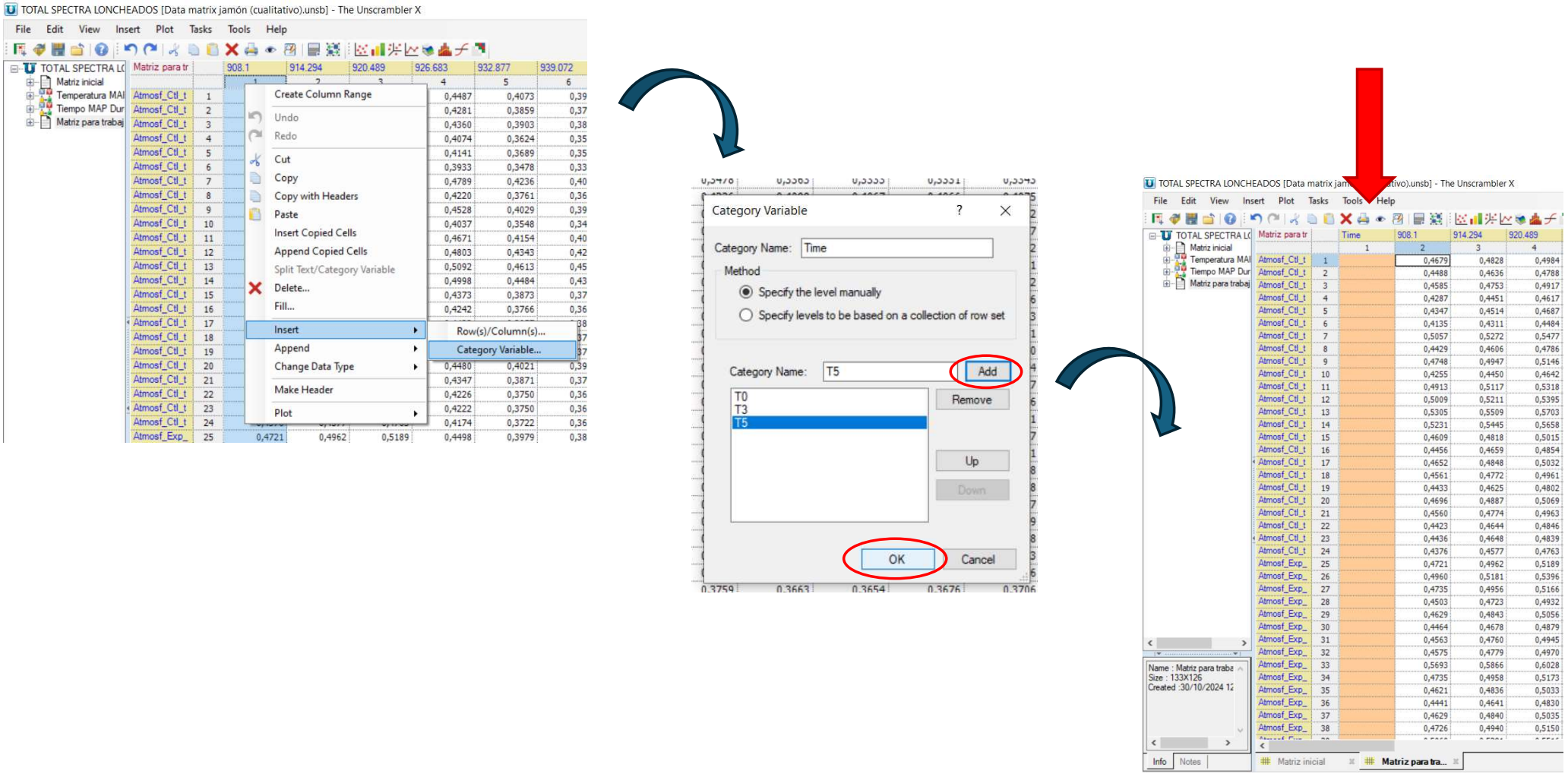
- Data import


TOTAL SPECTRA LONCHEADOS [Data matrix jamón (cualitativo).unsb] - The Unscrambler X
File Edit View Insert Plot Tasks Tools Help
TOTAL SPECTRA LC
Matriz inicial
Temperatura MAI
Tiempo MAP Dur
Matriz para trabaj
Name : Matriz para traba
Size : 133X125
Created :30/10/2024 12
Info Notes
Matriz para tr
908.1 914.294 920.489 926.683 932.877 939.072 945.266 951.46 957.655 963.849 970.043 976.238 982.432 988.626 994.821 1001.015 1007.209
Matriz inicial
Matriz para tra...
Ready
Value: 0,4679017
133X125

Qualitative models

- Data import
- **Insert row/column (to incorporate new samples or physical-chemical variables to predict)**
- **Insert Category variable**
- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)
- Development of classificatory models
- Predict

- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable



- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable

Assignment of dummy classes

01011
11010
01011

1 → Assigned to the sample belonging to the study class

0 → Assigned to the sample that does **NOT** belong to the study class.

Split Category Variable

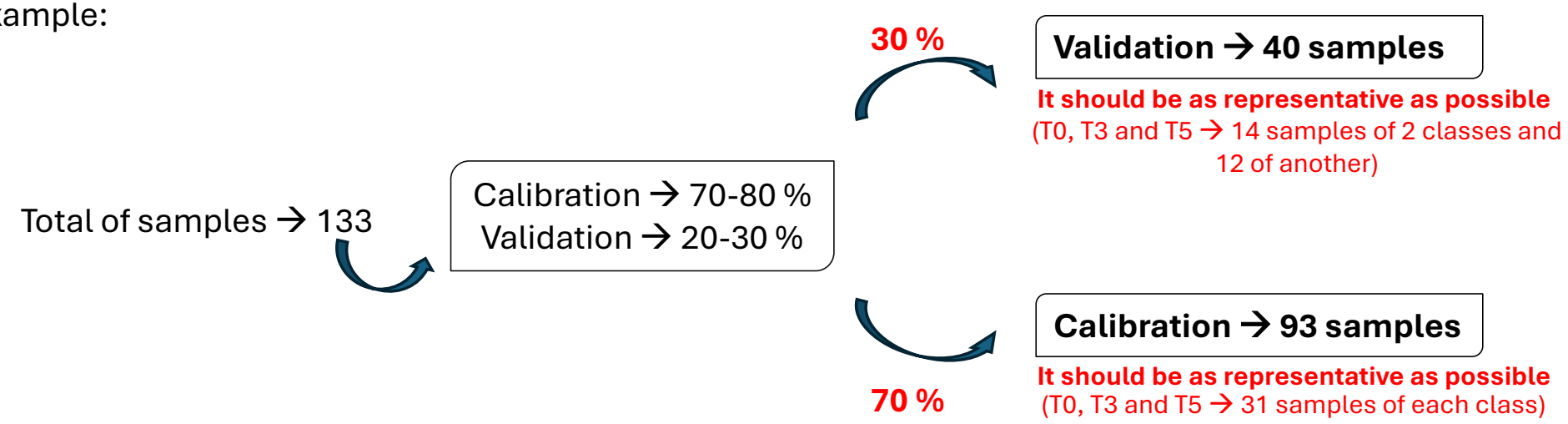
T0	T3	T5	Time
4	5	6	
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0
1,0000	0,0000	0,0000	T0

Qualitative models

- Data import
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- **Define range for the purpose of grouping rows and columns**
- **Sets (Calibration and Validation)**
- Development of classificatory models
- Predict

- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)

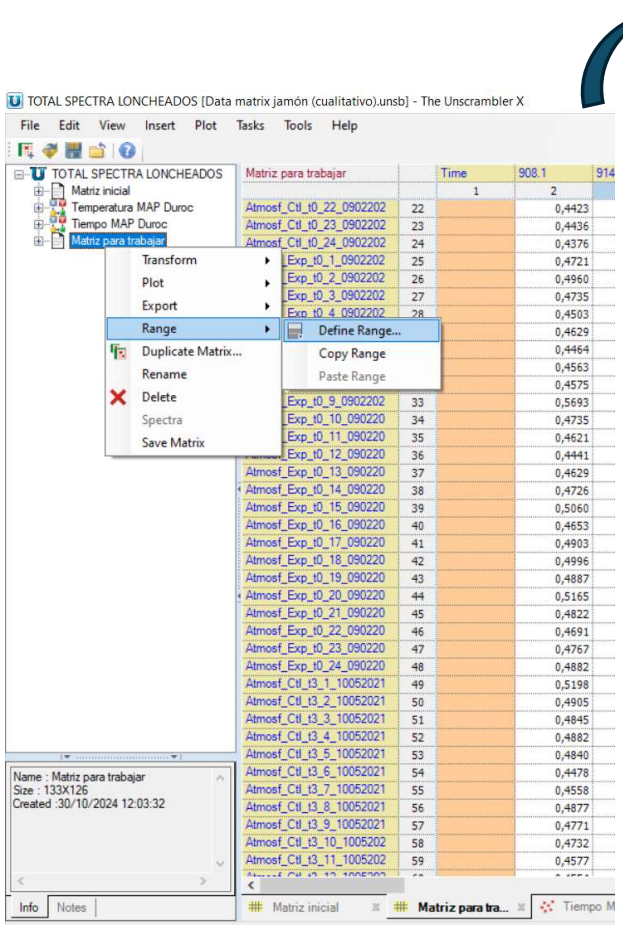
Example:

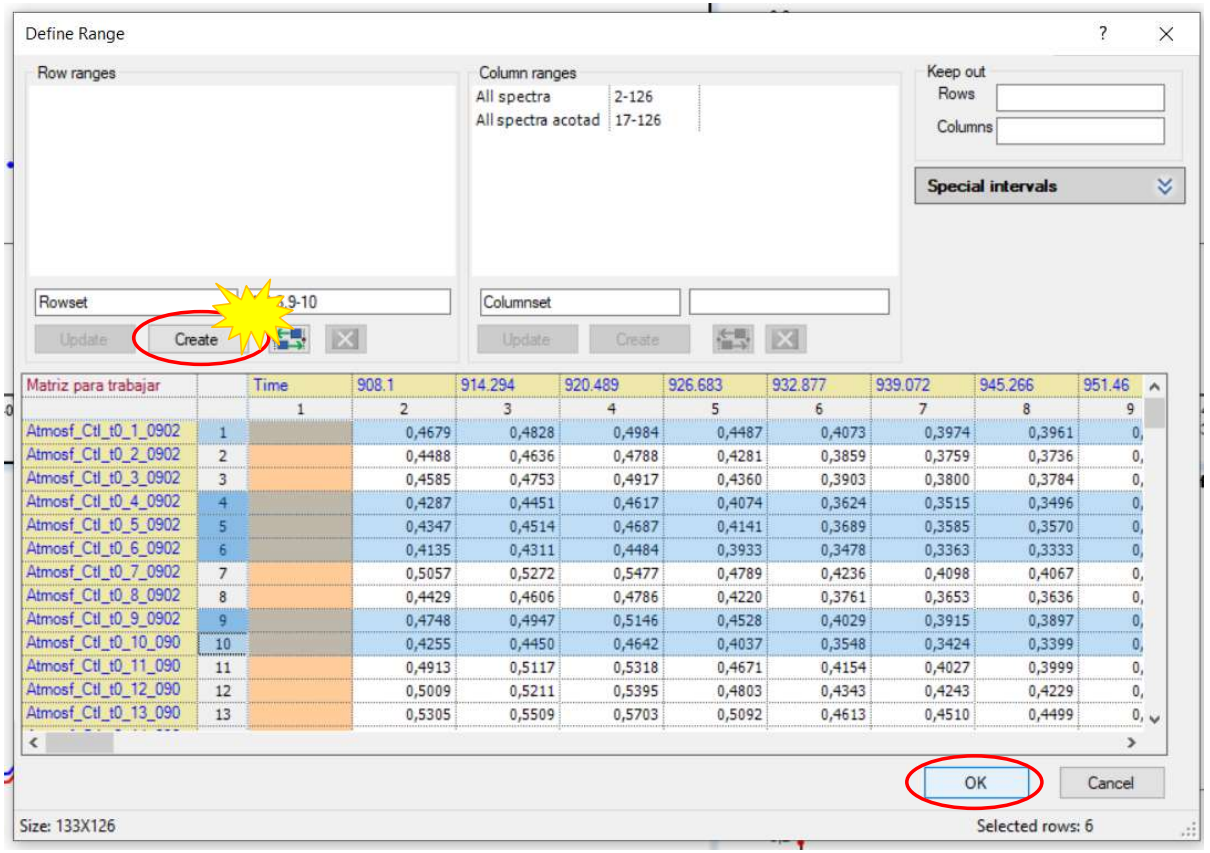


IMPORTANT!

Validation		
T0	T3	T5
14	14	12

- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)



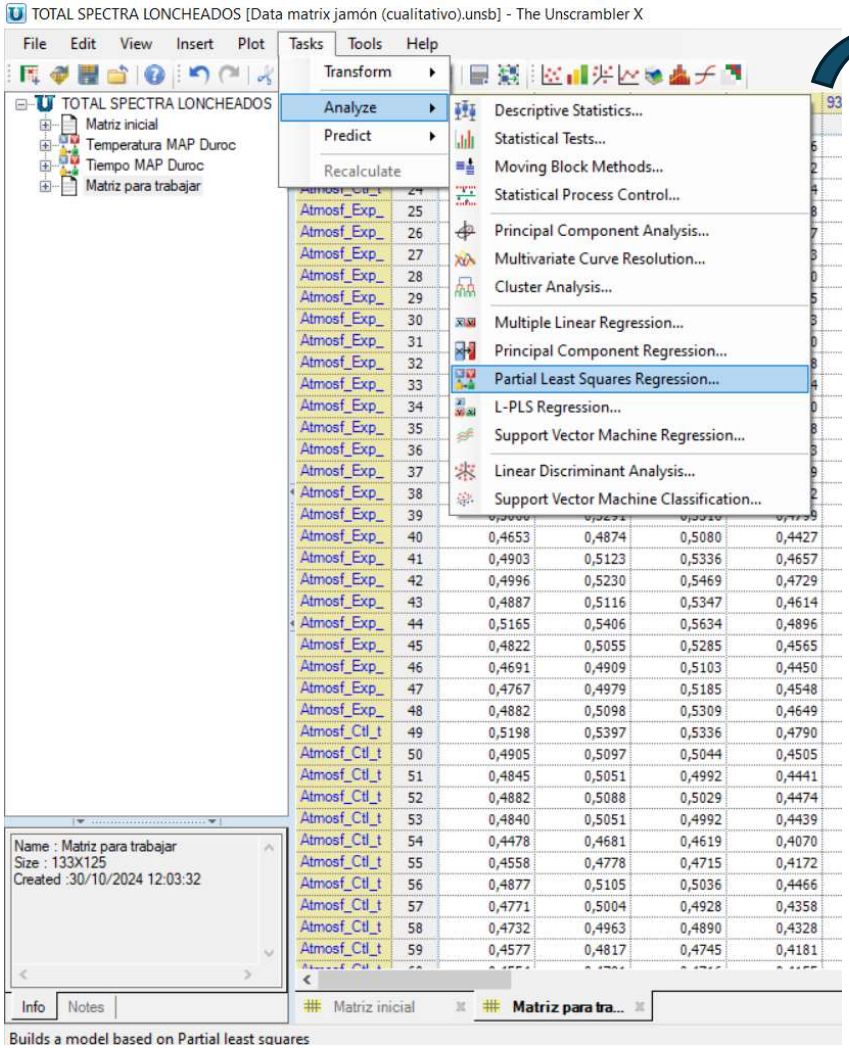


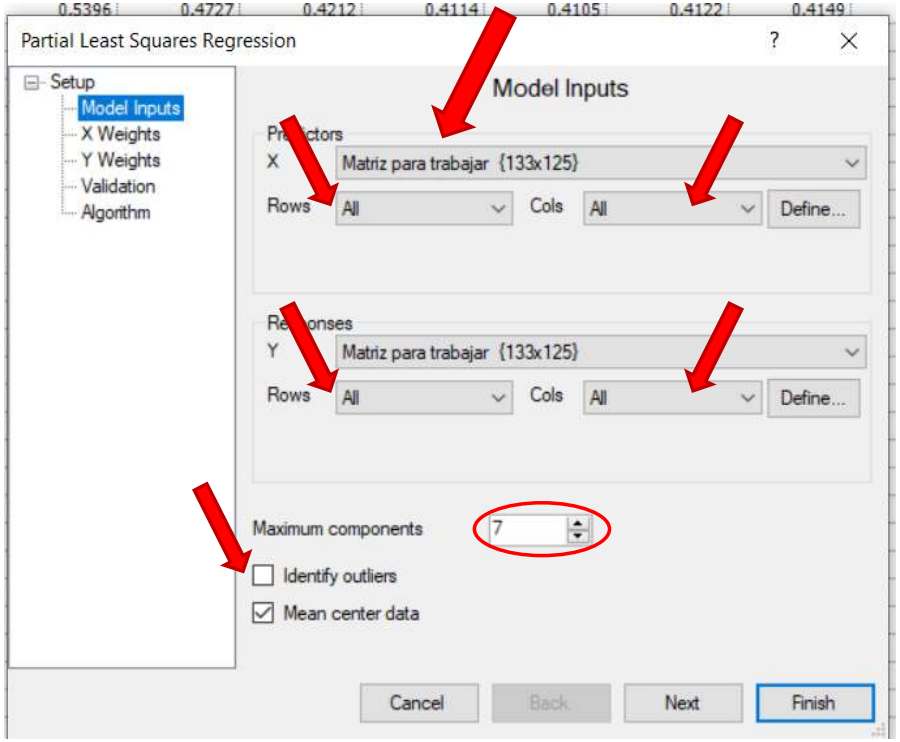
Matriz para trabajar		Time	908.1	914.294	920.489	926.683	932.877	939.072	945.266	951.46
		1	2	3	4	5	6	7	8	9
Atmosf_Ctl_t0_1_0902	1		0,4679	0,4828	0,4984	0,4487	0,4073	0,3974	0,3961	0,
Atmosf_Ctl_t0_2_0902	2		0,4488	0,4636	0,4788	0,4281	0,3859	0,3759	0,3736	0,
Atmosf_Ctl_t0_3_0902	3		0,4585	0,4753	0,4917	0,4360	0,3903	0,3800	0,3784	0,
Atmosf_Ctl_t0_4_0902	4		0,4287	0,4451	0,4617	0,4074	0,3624	0,3515	0,3496	0,
Atmosf_Ctl_t0_5_0902	5		0,4347	0,4514	0,4687	0,4141	0,3689	0,3585	0,3570	0,
Atmosf_Ctl_t0_6_0902	6		0,4135	0,4311	0,4484	0,3933	0,3478	0,3363	0,3333	0,
Atmosf_Ctl_t0_7_0902	7		0,5057	0,5272	0,5477	0,4789	0,4236	0,4098	0,4067	0,
Atmosf_Ctl_t0_8_0902	8		0,4429	0,4606	0,4786	0,4220	0,3761	0,3653	0,3636	0,
Atmosf_Ctl_t0_9_0902	9		0,4748	0,4947	0,5146	0,4528	0,4029	0,3915	0,3897	0,
Atmosf_Ctl_t0_10_090	10		0,4255	0,4450	0,4642	0,4037	0,3548	0,3424	0,3399	0,
Atmosf_Ctl_t0_11_090	11		0,4913	0,5117	0,5318	0,4671	0,4154	0,4027	0,3999	0,
Atmosf_Ctl_t0_12_090	12		0,5009	0,5211	0,5395	0,4803	0,4343	0,4243	0,4229	0,
Atmosf_Ctl_t0_13_090	13		0,5305	0,5509	0,5703	0,5092	0,4613	0,4510	0,4499	0,

Qualitative models

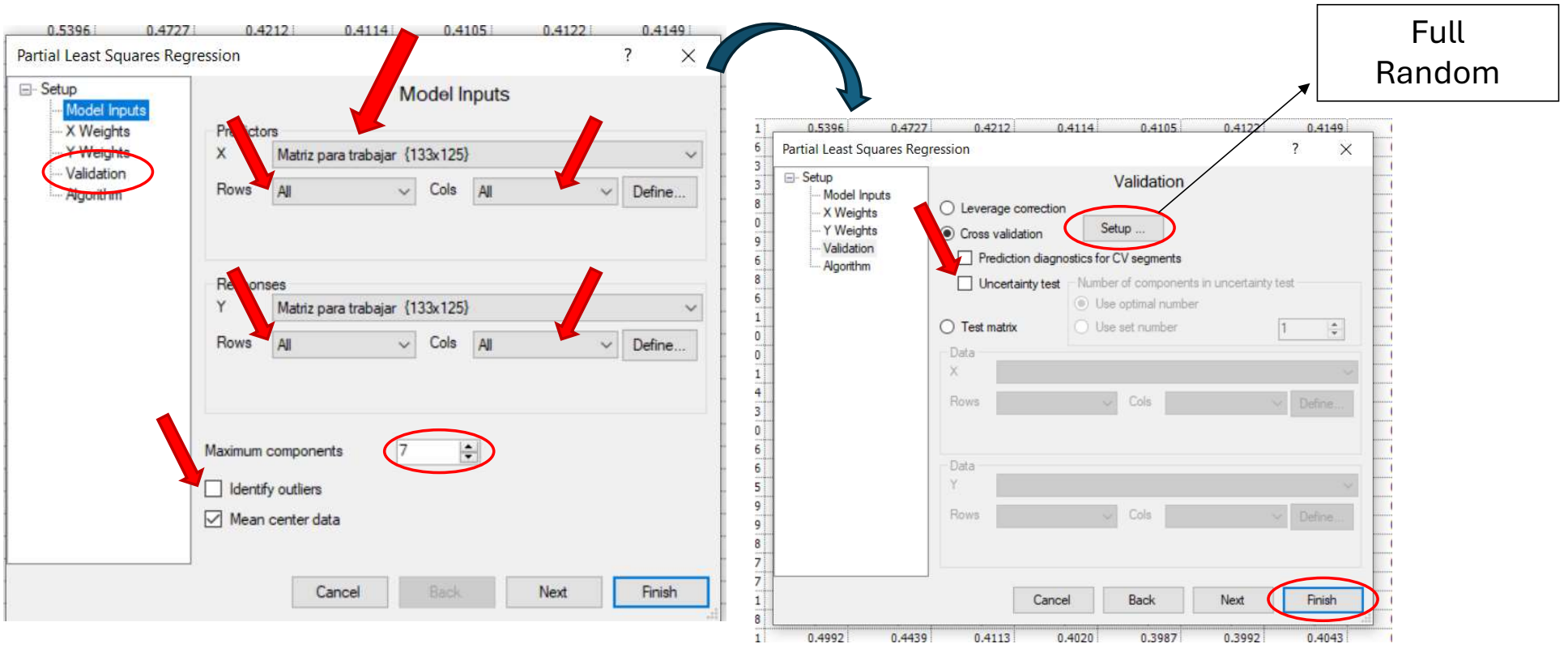
- Data import
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)
- **Development of classificatory models**
- Predict

- Development of classificatory models

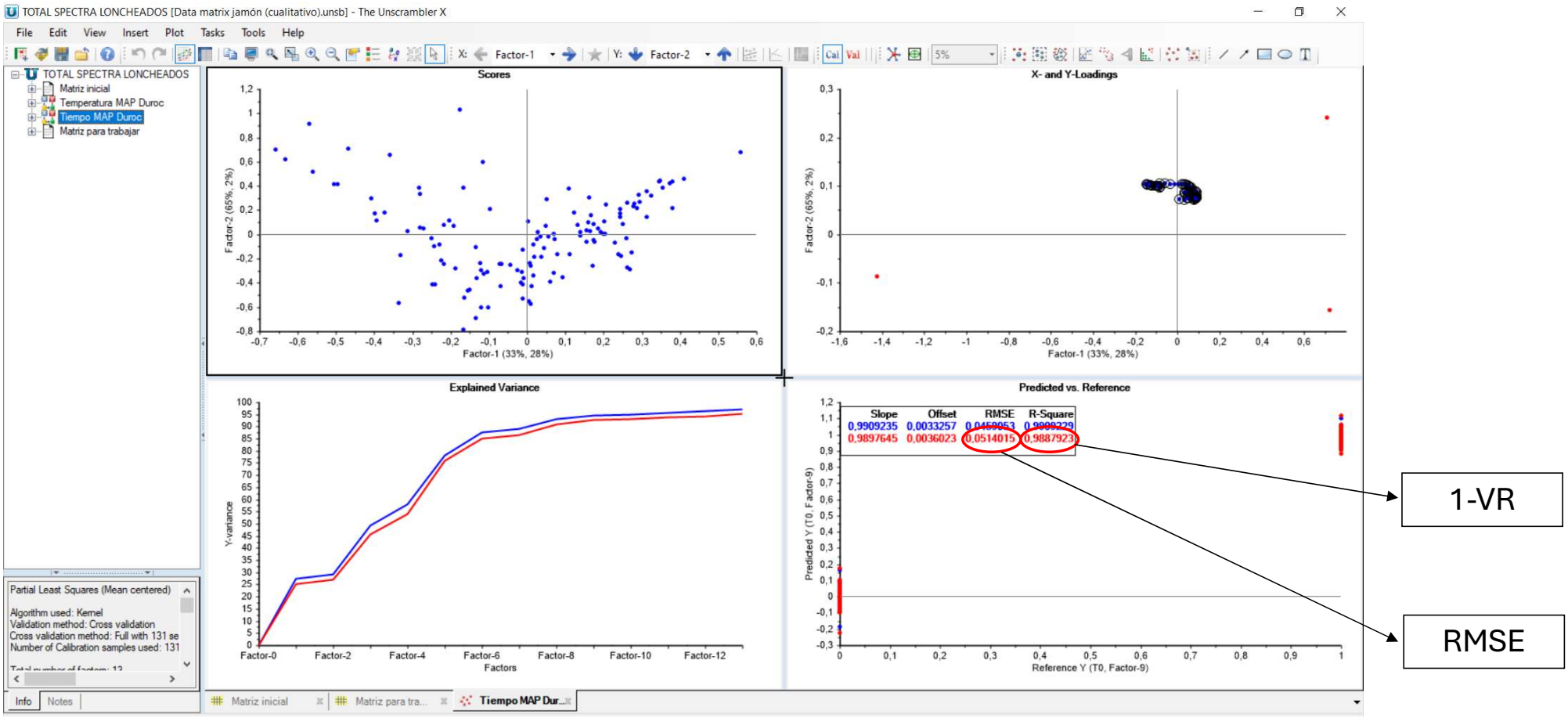




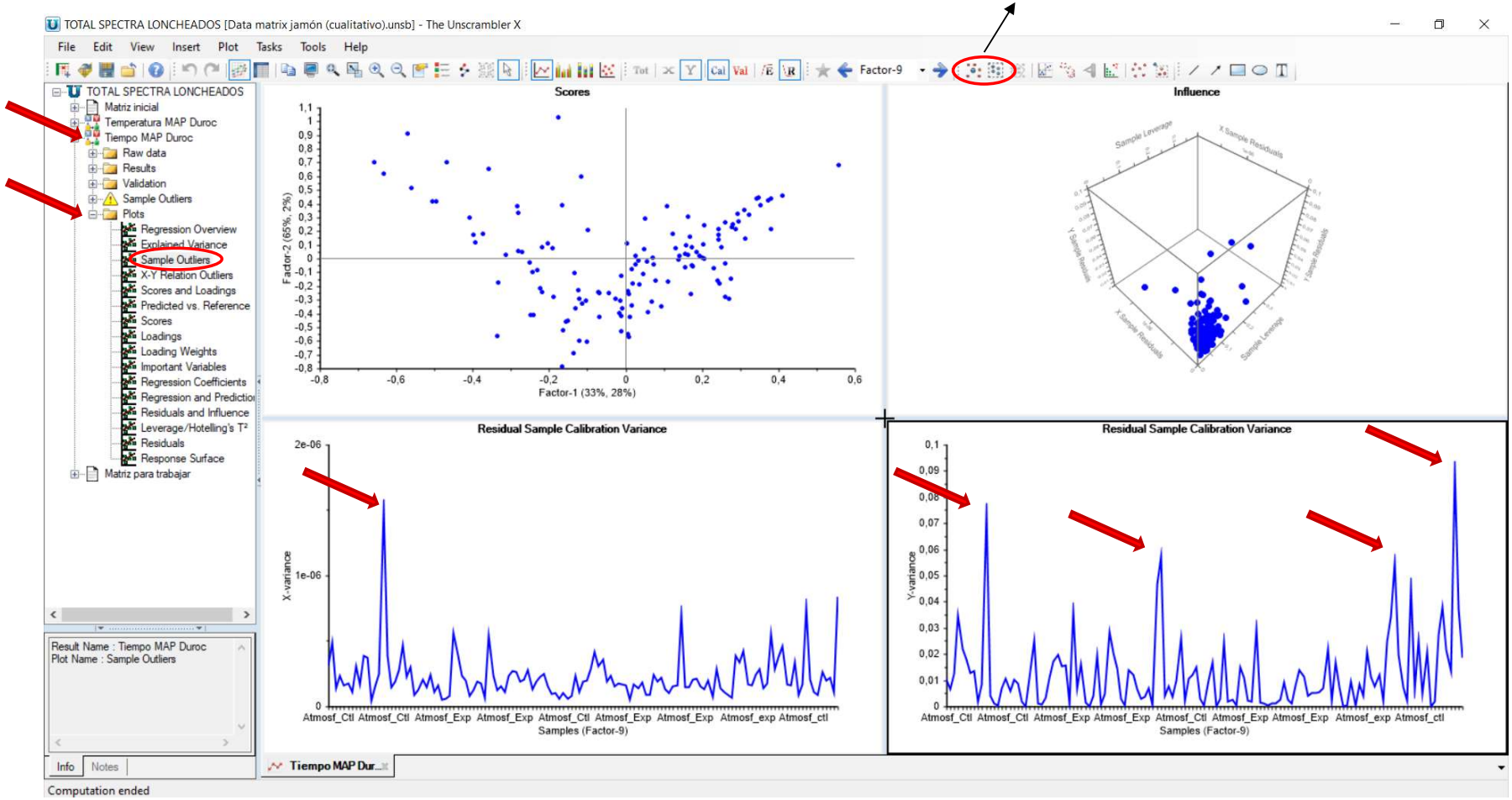
- Development of classificatory models



- Development of classificatory models



- Development of classificatory models



Qualitative models

- Data import
- Insert row/column (to incorporate new samples or physical-chemical variables to predict)
- Insert Category variable
- Define range for the purpose of grouping rows and columns
- Sets (Calibration and Validation)
- Development of classificatory models
- **Predict**

- Predict

STATISTICS

SENSITIVITY (SE)

Which denotes the percentage of samples belonging to an established class that the studied model has recognized as belonging to that class

$$SE\ (\%) = \frac{TP}{TP + FN}$$

TP = True positive
FN = False negative

SPECIFICITY (SP)

Which refers to the percentage of samples that do not belong to the target class and that the model has correctly rejected

$$SP\ (\%) = \frac{TN}{TN + FP}$$

TN = True negative
FP = False positive

- Predict

STATISTICS

ACCURACY

Defined as the percentage of samples correctly classified during external validation

$$Accuracy\ (\%) = \frac{TN + TP}{TN + TP + FN + FP}$$

TP = True positive
TN = True negative
FP = False positive
FN = False negative

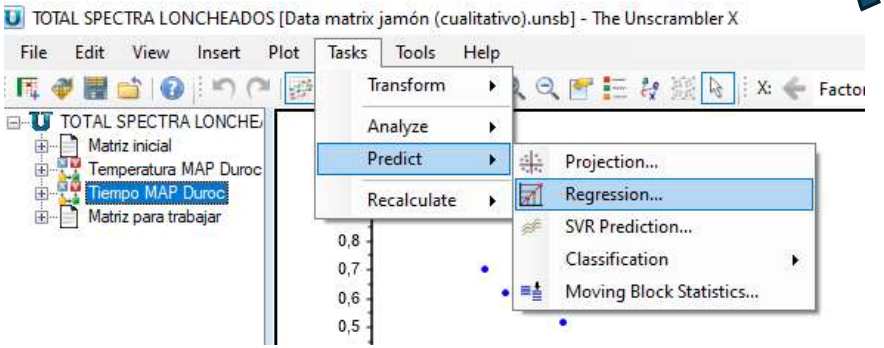
MATTHEWS CORRELATION COEFFICIENT (MCC)

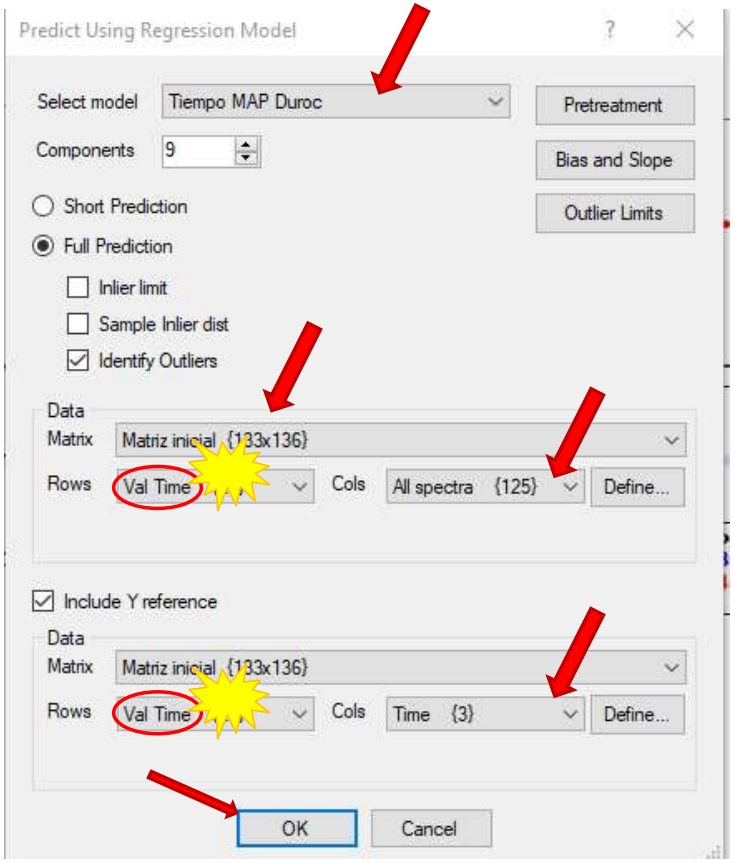
Which is suitable for assessing the performance of the model when the number of samples in the various classes is not the same

$$MCC = \frac{TP\ x\ TN - TP\ x\ FN}{\sqrt{(TP+FN)(TP+FN)(TN\ FN)(TN+TP)}}$$

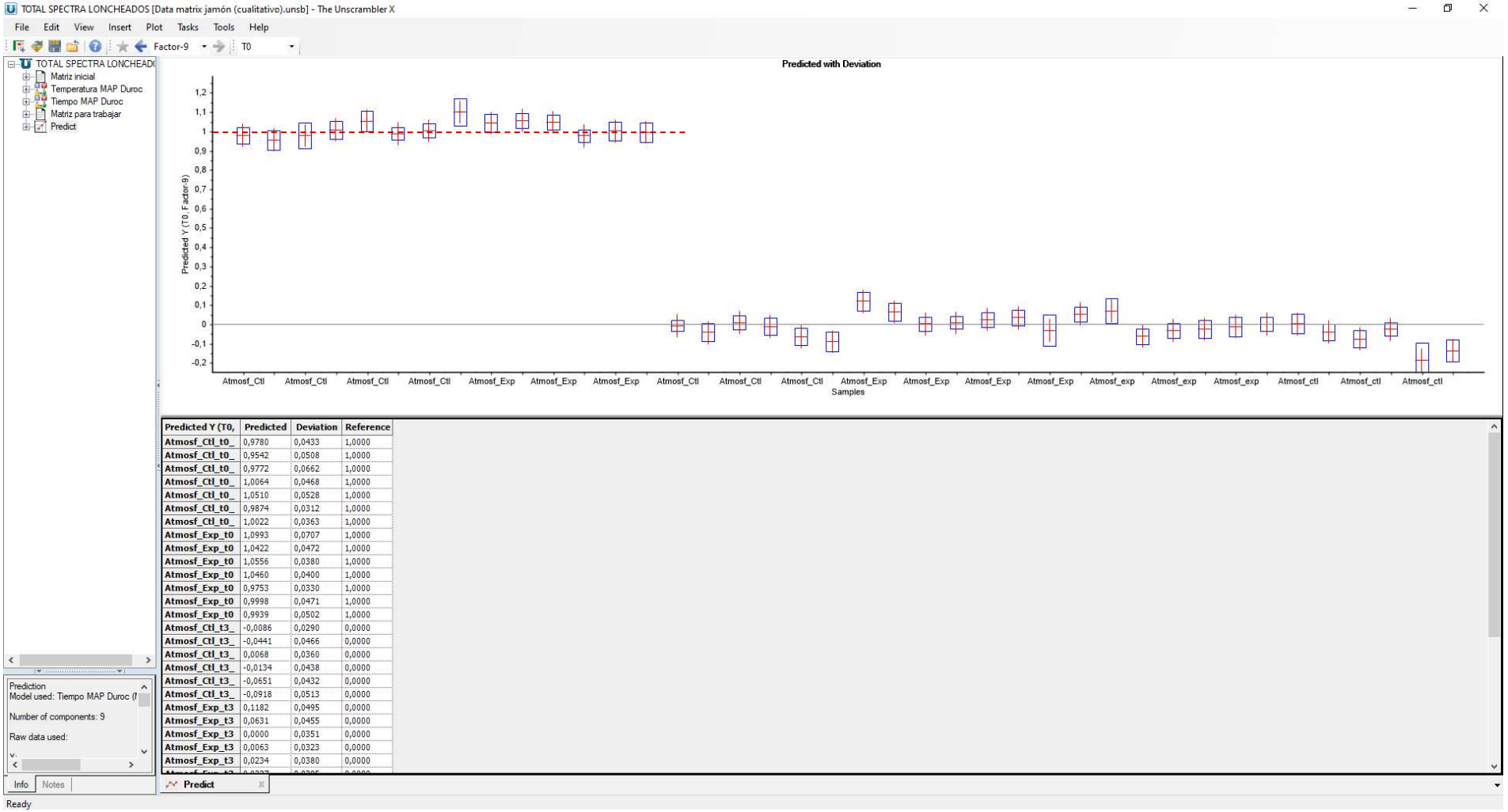
Value of 1 indicates a perfect prediction; -1 represents a total disagreement between the prediction and the actual values; and 0 means "no better than a random prediction"

- Predict


TOTAL SPECTRA LONCHEADOS [Data matrix jamón (cualitativo).unsb] - The Unscrambler X
File
Edit
View
Insert
Plot
Tasks
Tools
Help
TOTAL SPECTRA LONCHE
Matriz inicial
Temperatura MAP Duroc
Tiempo MAP Duroc
Matriz para trabajar
Transform
Analyze
Predict
Recalculate
Projection...
Regression...
SVR Prediction...
Classification
Moving Block Statistics...
0,8
0,7
0,6
0,5

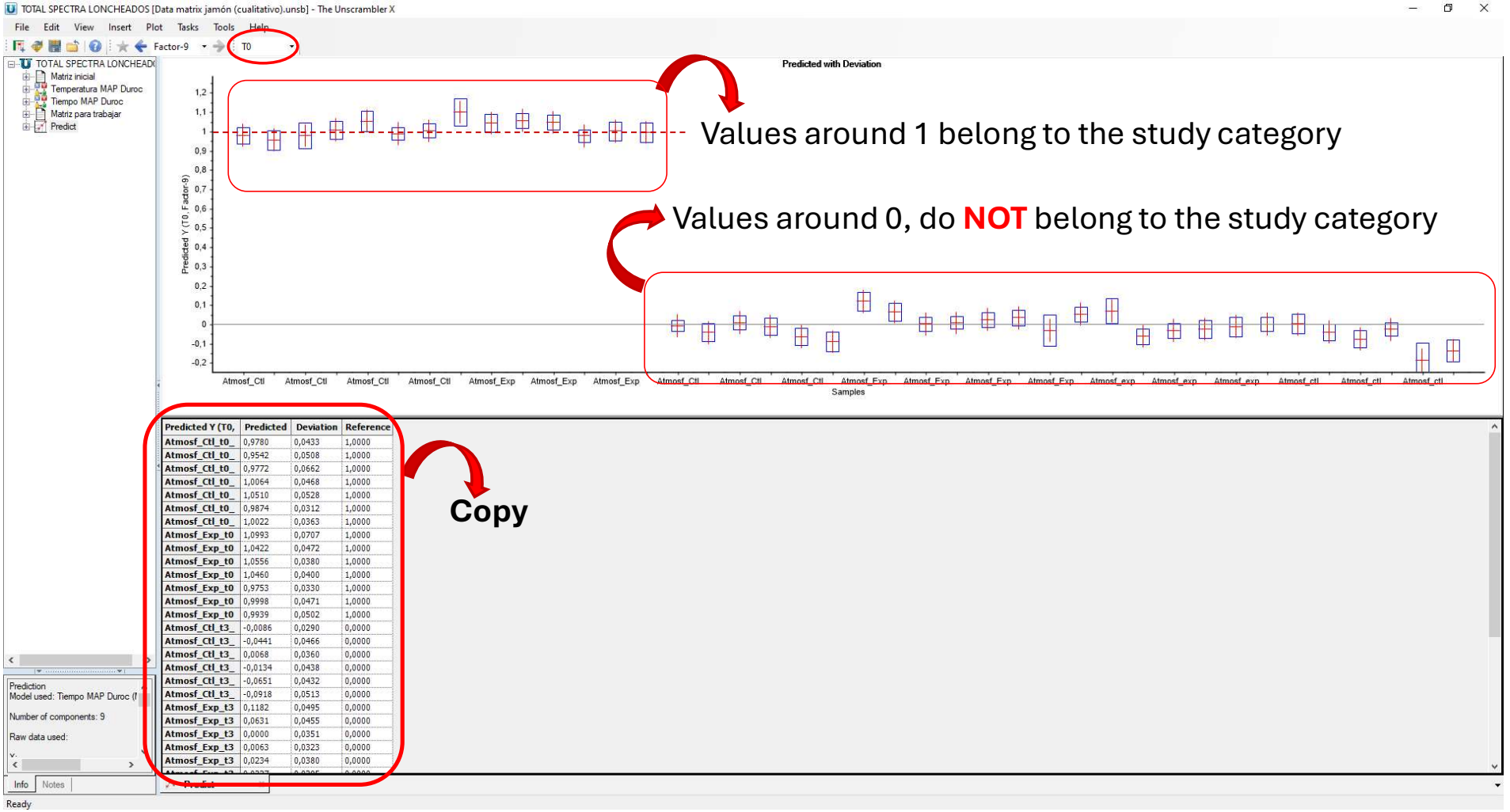

Predict Using Regression Model
Select model
Tiempo MAP Duroc
Pretreatment
Components
9
Bias and Slope
Short Prediction
Outlier Limits
Full Prediction
Inlier limit
Sample Inlier dist
Identify Outliers
Data
Matrix
Matriz inicial {133x136}
Rows
Val Time
Cols
All spectra {125}
Define...
Include Y reference
Data
Matrix
Matriz inicial {133x136}
Rows
Val Time
Cols
Time {3}
Define...
OK
Cancel

- Predict



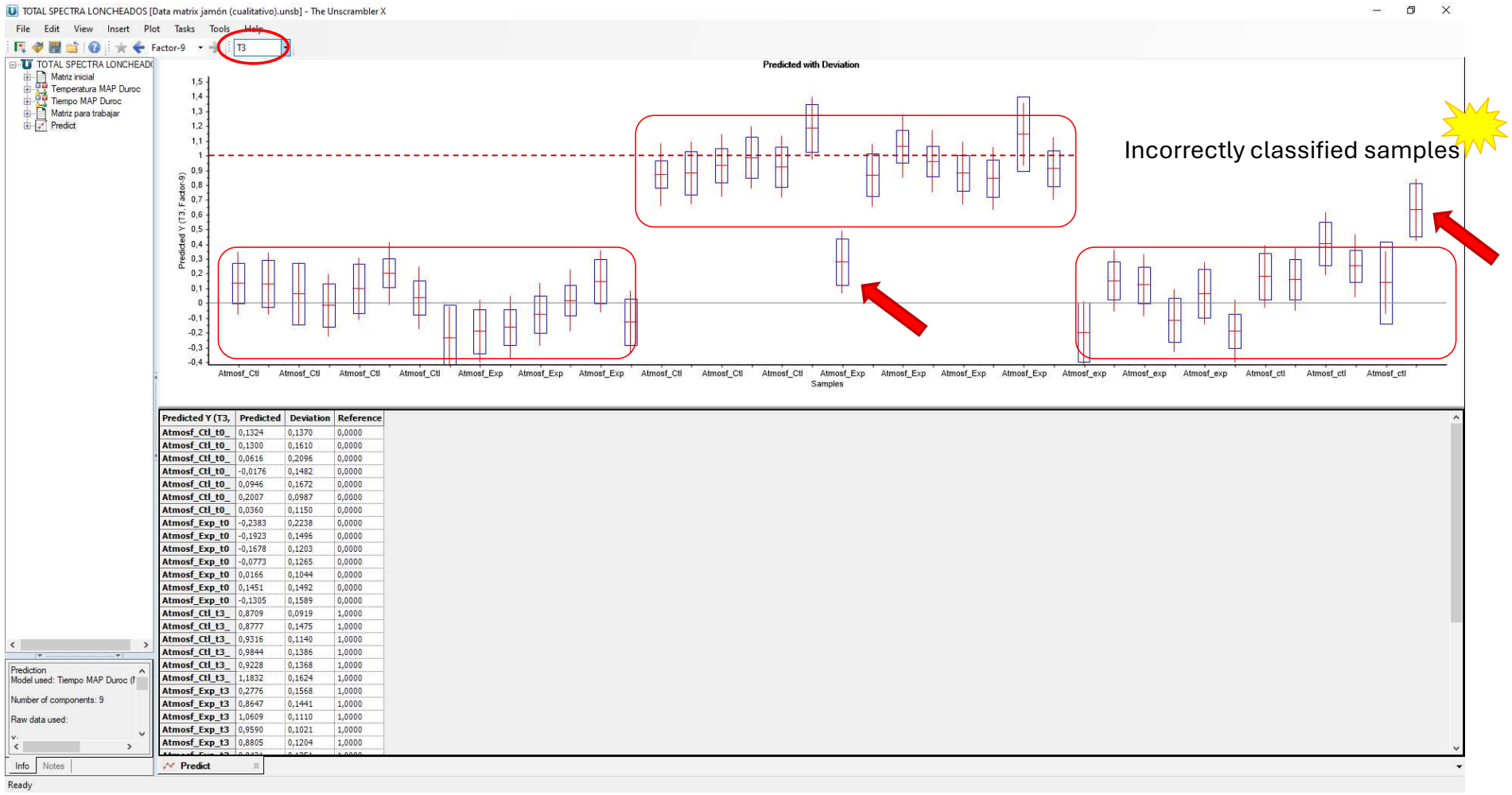
Predicted Y (T0,	Predicted	Deviation	Reference
Atmosf_Ctl_t0_	0,9780	0,0433	1,0000
Atmosf_Ctl_t0_	0,9542	0,0508	1,0000
Atmosf_Ctl_t0_	0,9772	0,0662	1,0000
Atmosf_Ctl_t0_	1,0064	0,0468	1,0000
Atmosf_Ctl_t0_	1,0510	0,0528	1,0000
Atmosf_Ctl_t0_	0,9874	0,0312	1,0000
Atmosf_Ctl_t0_	1,0022	0,0363	1,0000
Atmosf_Exp_t0	1,0993	0,0707	1,0000
Atmosf_Exp_t0	1,0422	0,0472	1,0000
Atmosf_Exp_t0	1,0556	0,0380	1,0000
Atmosf_Exp_t0	1,0460	0,0400	1,0000
Atmosf_Exp_t0	0,9753	0,0330	1,0000
Atmosf_Exp_t0	0,9998	0,0471	1,0000
Atmosf_Exp_t0	0,9939	0,0502	1,0000
Atmosf_Ctl_t3_	-0,0086	0,0290	0,0000
Atmosf_Ctl_t3_	-0,0441	0,0466	0,0000
Atmosf_Ctl_t3_	0,0068	0,0360	0,0000
Atmosf_Ctl_t3_	-0,0134	0,0438	0,0000
Atmosf_Ctl_t3_	-0,0651	0,0432	0,0000
Atmosf_Ctl_t3_	-0,0918	0,0513	0,0000
Atmosf_Exp_t3	0,1182	0,0495	0,0000
Atmosf_Exp_t3	0,0631	0,0455	0,0000
Atmosf_Exp_t3	0,0000	0,0351	0,0000
Atmosf_Exp_t3	0,0063	0,0323	0,0000
Atmosf_Exp_t3	0,0234	0,0380	0,0000

- Predict


TOTAL SPECTRA LONCHEADOS [Data matrix jamón (cualitativo).unsb] - The Unscrambler X
File Edit View Insert Plot Tasks Tools Help
Factor-9
T0
TOTAL SPECTRA LONCHEADO
Matriz inicial
Temperatura MAP Duroc
Tiempo MAP Duroc
Matriz para trabajar
Predict
Predicted with Deviation
Predicted Y (T0, Factor-9)
Samples
Values around 1 belong to the study category
Values around 0, do NOT belong to the study category
Copy
Predicted Y (T0, | Predicted | Deviation | Reference
Atmosf_Ctl_t0_ | 0,9780 | 0,0433 | 1,0000
Atmosf_Ctl_t0_ | 0,9542 | 0,0508 | 1,0000
Atmosf_Ctl_t0_ | 0,9772 | 0,0662 | 1,0000
Atmosf_Ctl_t0_ | 1,0064 | 0,0468 | 1,0000
Atmosf_Ctl_t0_ | 1,0510 | 0,0528 | 1,0000
Atmosf_Ctl_t0_ | 0,9874 | 0,0312 | 1,0000
Atmosf_Ctl_t0_ | 1,0022 | 0,0363 | 1,0000
Atmosf_Exp_t0 | 1,0993 | 0,0707 | 1,0000
Atmosf_Exp_t0 | 1,0422 | 0,0472 | 1,0000
Atmosf_Exp_t0 | 1,0556 | 0,0380 | 1,0000
Atmosf_Exp_t0 | 1,0460 | 0,0400 | 1,0000
Atmosf_Exp_t0 | 0,9753 | 0,0330 | 1,0000
Atmosf_Exp_t0 | 0,9998 | 0,0471 | 1,0000
Atmosf_Exp_t0 | 0,9939 | 0,0502 | 1,0000
Atmosf_Ctl_t3_ | -0,0086 | 0,0290 | 0,0000
Atmosf_Ctl_t3_ | -0,0441 | 0,0466 | 0,0000
Atmosf_Ctl_t3_ | 0,0068 | 0,0360 | 0,0000
Atmosf_Ctl_t3_ | -0,0134 | 0,0438 | 0,0000
Atmosf_Ctl_t3_ | -0,0651 | 0,0432 | 0,0000
Atmosf_Ctl_t3_ | -0,0918 | 0,0513 | 0,0000
Atmosf_Exp_t3 | 0,1182 | 0,0495 | 0,0000
Atmosf_Exp_t3 | 0,0631 | 0,0455 | 0,0000
Atmosf_Exp_t3 | 0,0000 | 0,0351 | 0,0000
Atmosf_Exp_t3 | 0,0063 | 0,0323 | 0,0000
Atmosf_Exp_t3 | 0,0234 | 0,0380 | 0,0000
Prediction
Model used: Tiempo MAP Duroc (
Number of components: 9
Raw data used:
Info
Notes
Ready

- Predict

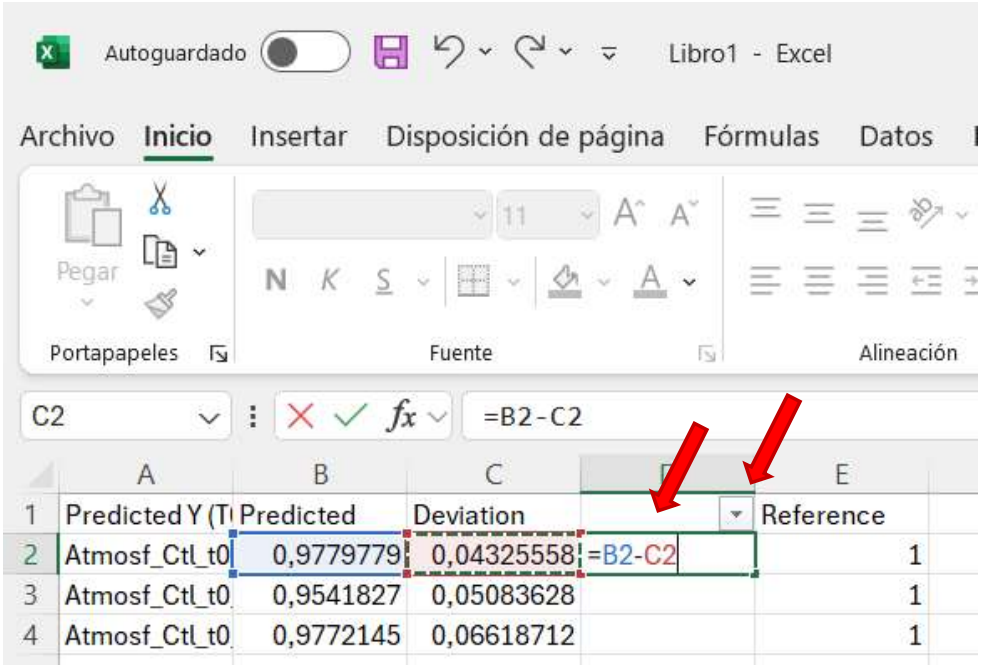
  Another example:



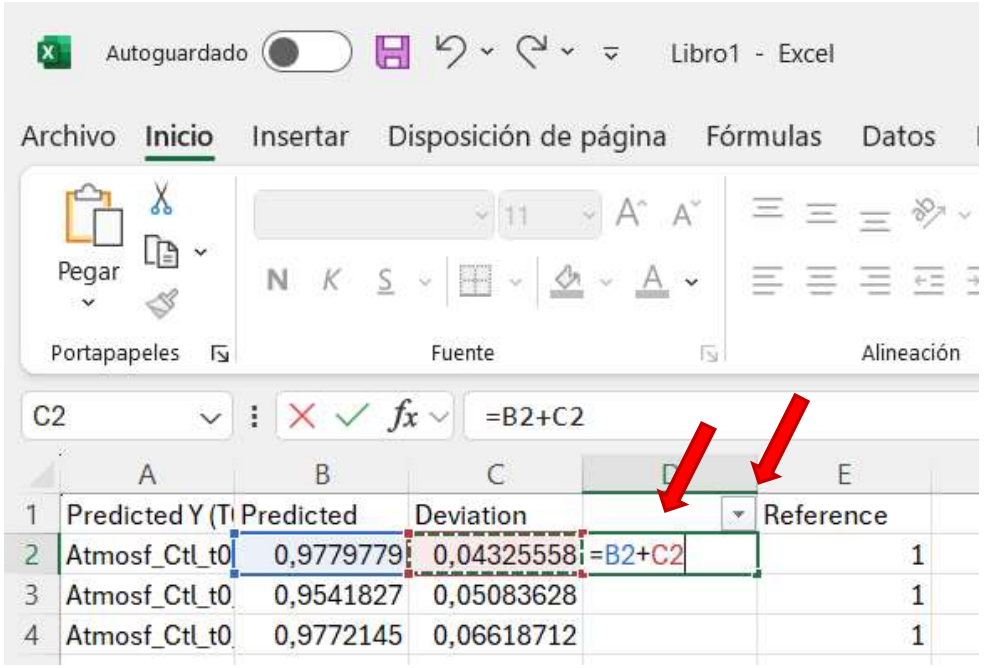
- Predict

Predicted value - deviation ≥ 0.5 → **SENSITIVITY**.
Predicted value + deviation < 0.5 → **SPECIFICITY**.

We calculate the most unfavorable case



SENSITIVITY



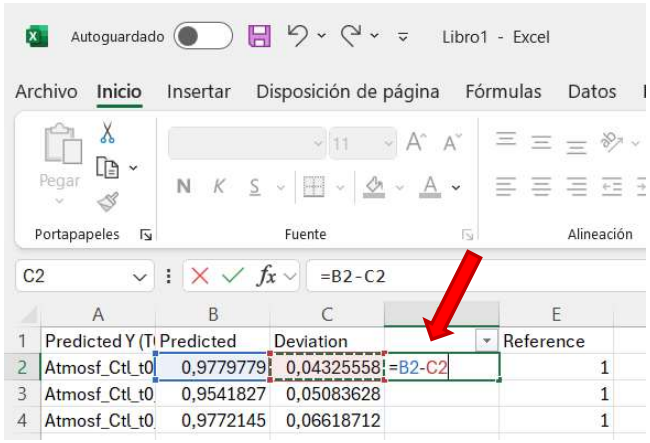
SPECIFICITY

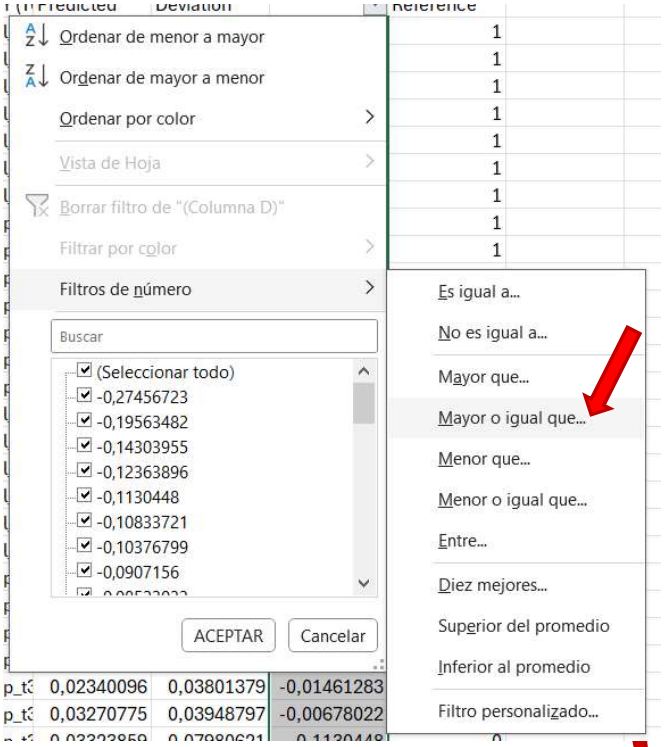
- Predict

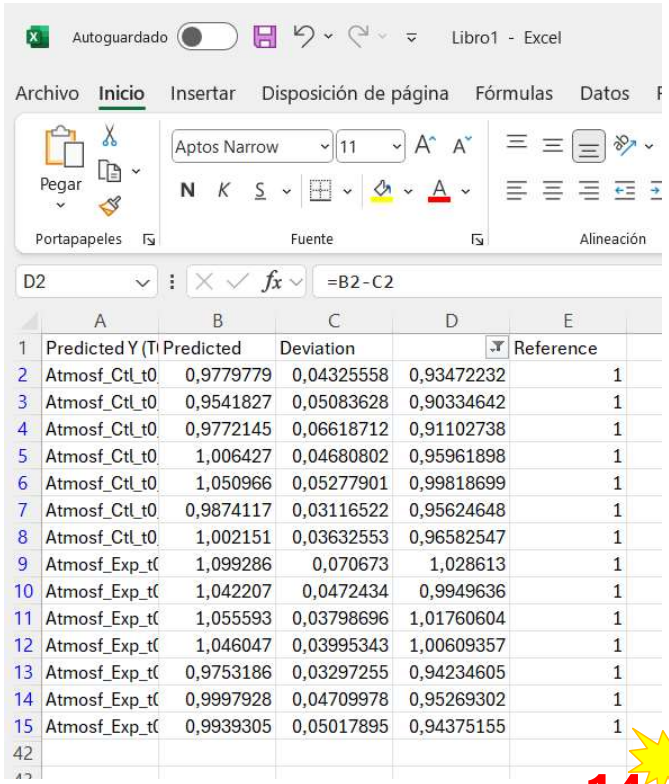
Predicted value - deviation ≥ 0.5 → **SENSITIVITY**.
Predicted value + deviation < 0.5 → **SPECIFICITY**.

We calculate the most unfavorable case

SENSITIVITY







14
True positive

- Predict

Validation		
T0	T3	T5
14	14	12

14
True positive

$$SE\ (\%) = \frac{TP}{TP + FN}$$

$$SE\ (\%) = \frac{14}{14 + 0}$$

100%

TP = True positive
FN = False negative

FN = Total class samples - TP

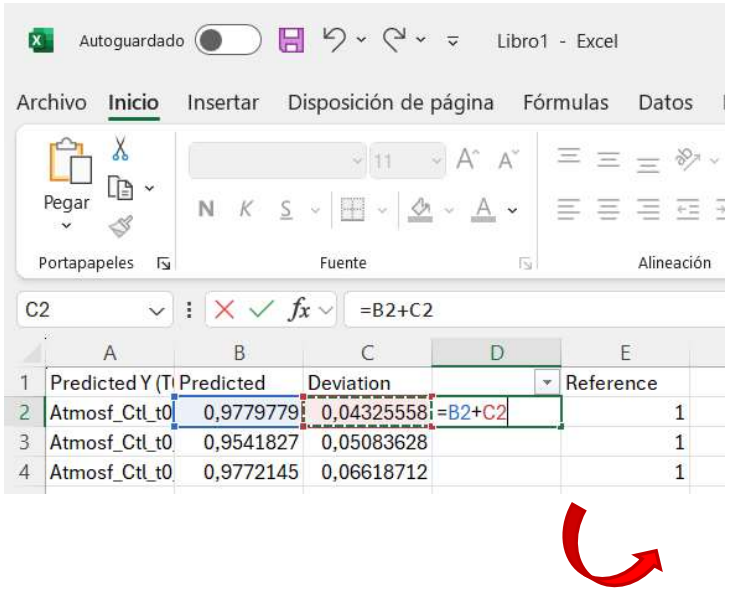
TP = 14
FN = 0

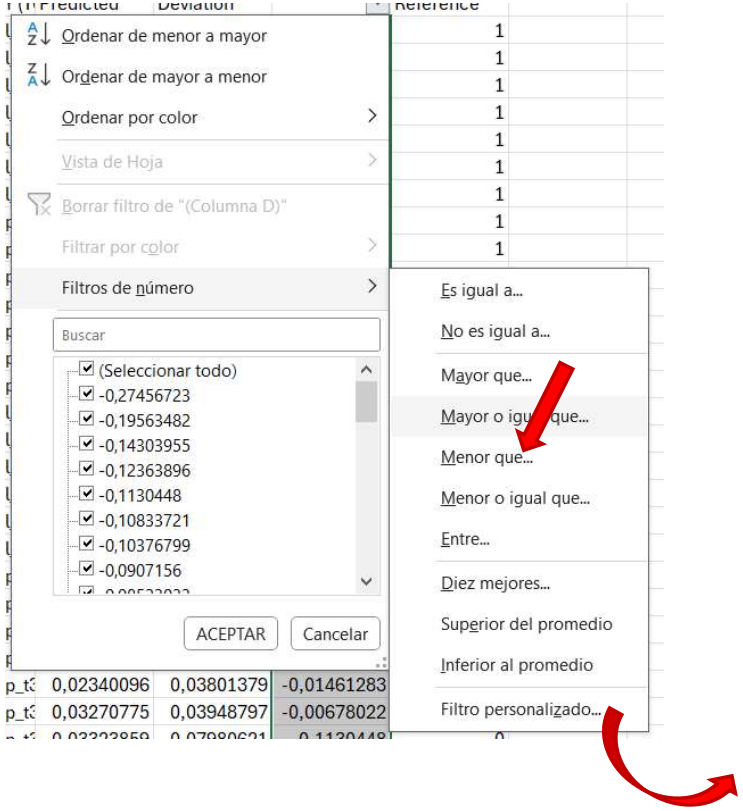
- Predict

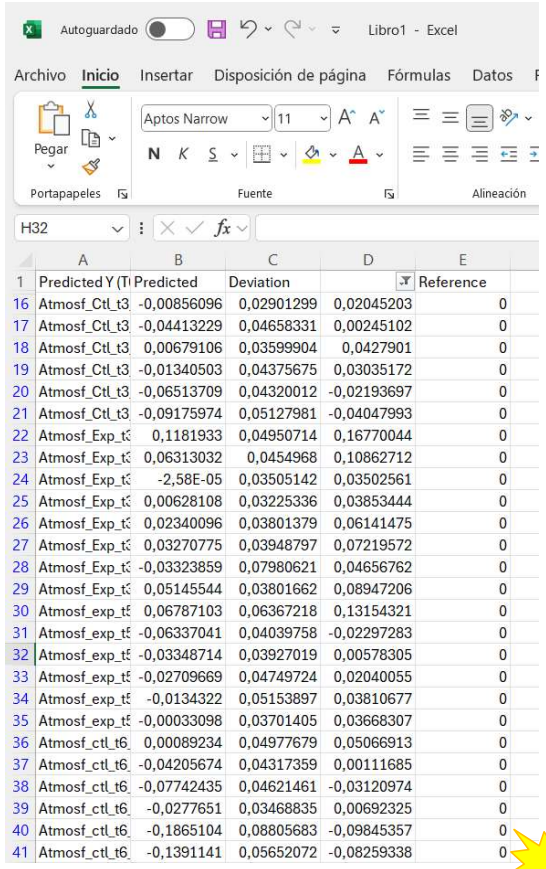
Predicted value - deviation ≥ 0.5 → **SENSITIVITY.**
Predicted value + deviation < 0.5 → **SPECIFICITY.**

We calculate the most unfavorable case

SPECIFICITY

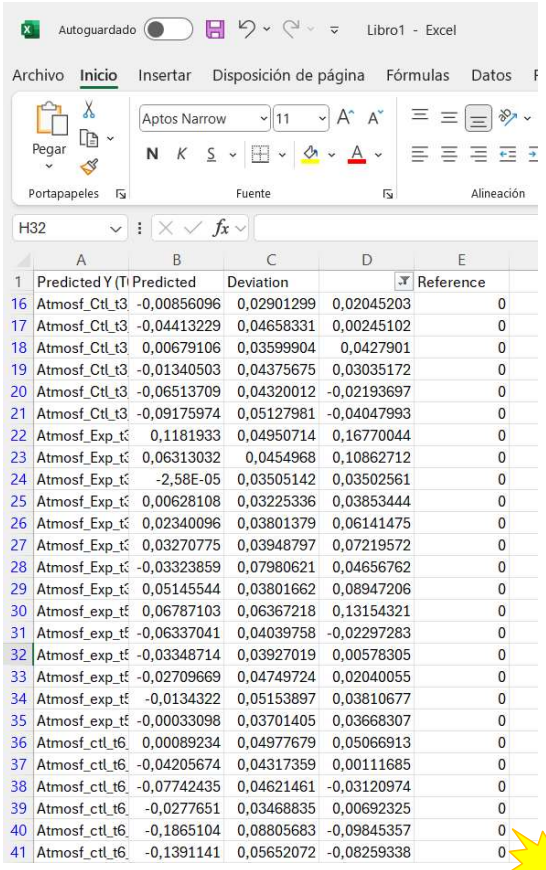






26
True negative

- Predict



	A	B	C	D	E
1	Predicted Y (T	Predicted	Deviation		Reference
16	Atmosf_Ctl_t3	-0,00856096	0,02901299	0,02045203	0
17	Atmosf_Ctl_t3	-0,04413229	0,04658331	0,00245102	0
18	Atmosf_Ctl_t3	0,00679106	0,03599904	0,0427901	0
19	Atmosf_Ctl_t3	-0,01340503	0,04375675	0,03035172	0
20	Atmosf_Ctl_t3	-0,06513709	0,04320012	-0,02193697	0
21	Atmosf_Ctl_t3	-0,09175974	0,05127981	-0,04047993	0
22	Atmosf_Exp_t3	0,1181933	0,04950714	0,16770044	0
23	Atmosf_Exp_t3	0,06313032	0,0454968	0,10862712	0
24	Atmosf_Exp_t3	-2,58E-05	0,03505142	0,03502561	0
25	Atmosf_Exp_t3	0,00628108	0,03225336	0,03853444	0
26	Atmosf_Exp_t3	0,02340096	0,03801379	0,06141475	0
27	Atmosf_Exp_t3	0,03270775	0,03948797	0,07219572	0
28	Atmosf_Exp_t3	-0,03323859	0,07980621	0,04656762	0
29	Atmosf_Exp_t3	0,05145544	0,03801662	0,08947206	0
30	Atmosf_exp_t5	0,06787103	0,06367218	0,13154321	0
31	Atmosf_exp_t5	-0,06337041	0,04039758	-0,02297283	0
32	Atmosf_exp_t5	-0,03348714	0,03927019	0,00578305	0
33	Atmosf_exp_t5	-0,02709669	0,04749724	0,02040055	0
34	Atmosf_exp_t5	-0,0134322	0,05153897	0,03810677	0
35	Atmosf_exp_t5	-0,00033098	0,03701405	0,03668307	0
36	Atmosf_ctl_t6_	0,00089234	0,04977679	0,05066913	0
37	Atmosf_ctl_t6_	-0,04205674	0,04317359	0,00111685	0
38	Atmosf_ctl_t6_	-0,07742435	0,04621461	-0,03120974	0
39	Atmosf_ctl_t6_	-0,0277651	0,03468835	0,00692325	0
40	Atmosf_ctl_t6_	-0,1865104	0,08805683	-0,09845357	0
41	Atmosf_ctl_t6_	-0,1391141	0,05652072	-0,08259338	0

26

True negative

Validation		
T0	T3	T5
14	14	12

26

$$SP\ (\%) = \frac{TN}{TN + FP}$$

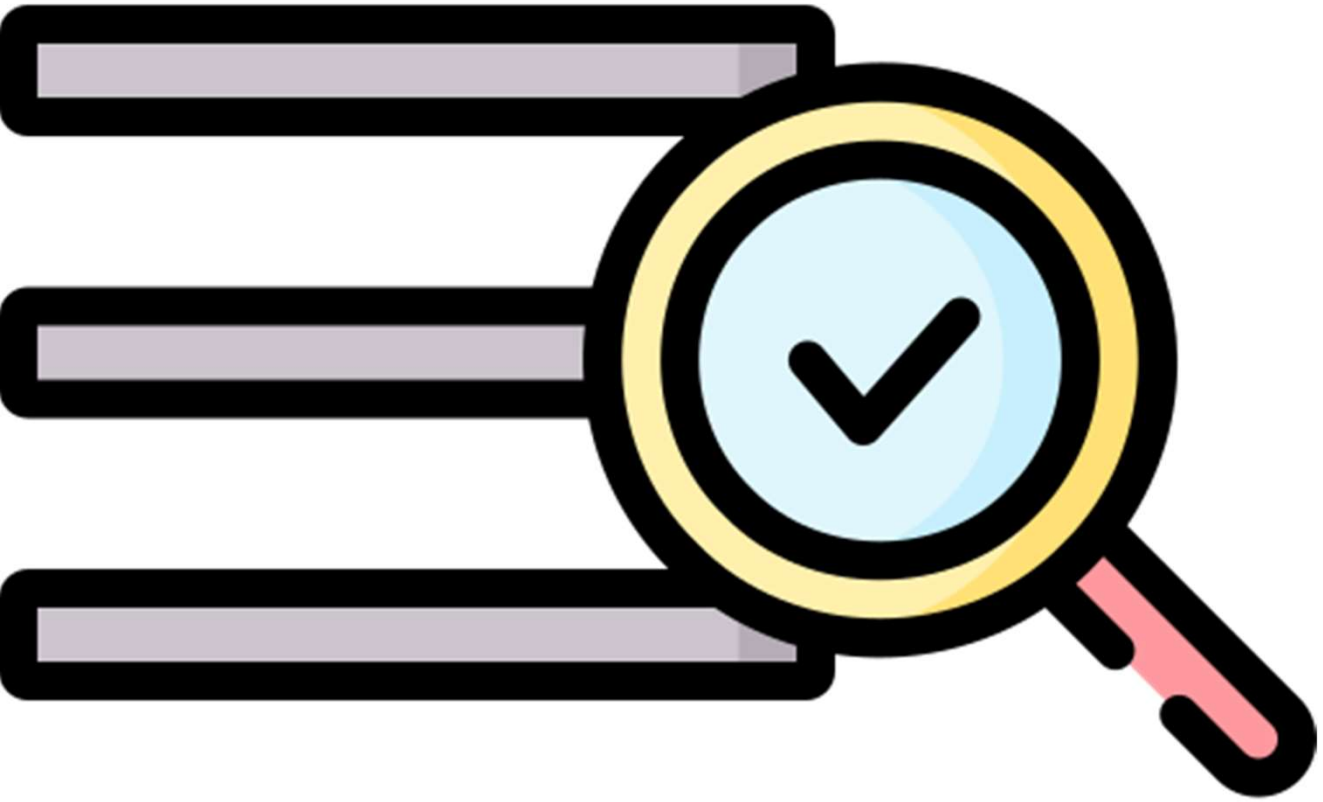
$$SP\ (\%) = \frac{26}{26 + 0}$$

100%

TN = True negative
FP = False positive

TN = 26
FP = 0

FP = Total of samples not belonging to the class - TN



Index

- What is NIRS technology?
- The Unscrambler
  - The construction of the matrix
  - Quantitative models
  - Qualitative models
  - **Pre-treatments**

- Pre-treatments

Los **efectos no deseados presentes en las señales** obtenidas en espectroscopia se denominan comúnmente **ruido**.

Este ruido puede tener diferentes causas u orígenes.

- ❑ Puede haber ruido debido a los componentes de la instrumentación utilizada para el registro del espectro (ruido instrumental) (variaciones de temperatura, humedad u otras)
- ❑ o bien variaciones en la señal debidas a la propia naturaleza de la muestra

Pre-treatments

Tratamientos espectrales para reducir el ruido

- Pre-treatments

❑ **Scattering effect** refers to the **dispersion of incident radiation** due to **physical phenomena** (texture, size and geometry of the particles that make up the sample). This causes unwanted variations in the spectral data.

<u>Scatter correction treatments</u>

❑ **Detrending (DT)**
❑ **Standard Normal Variate (SNV)**
❑ **Multiplicative Scatter Correction (MSC)**

❑ **Average**

Averaging several analytical signals of the same sample decreases noise and improves the signal-to-noise ratio.

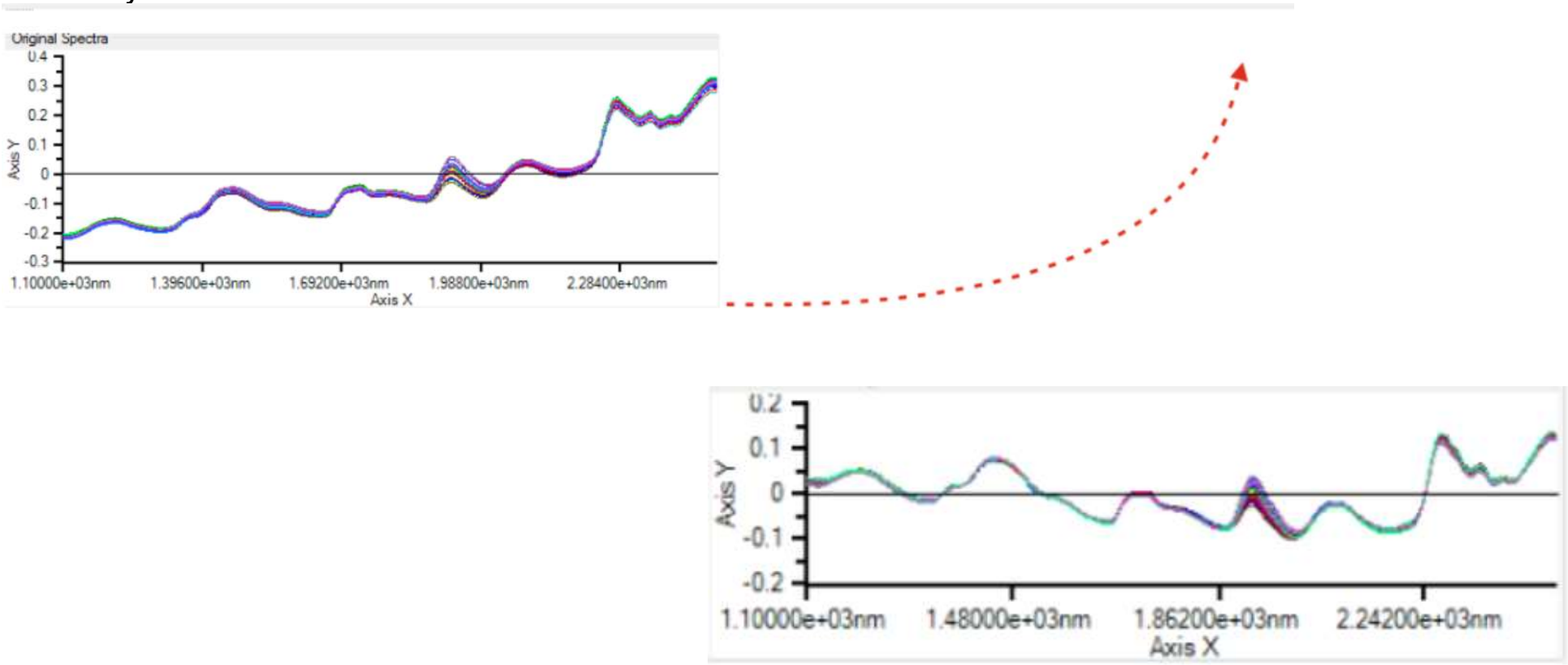
❑ **Spectral smoothing** to **mathematically** reduce random noise and increase signal-to-noise ratio

❑ **Derivatives**

- Pre-treatments

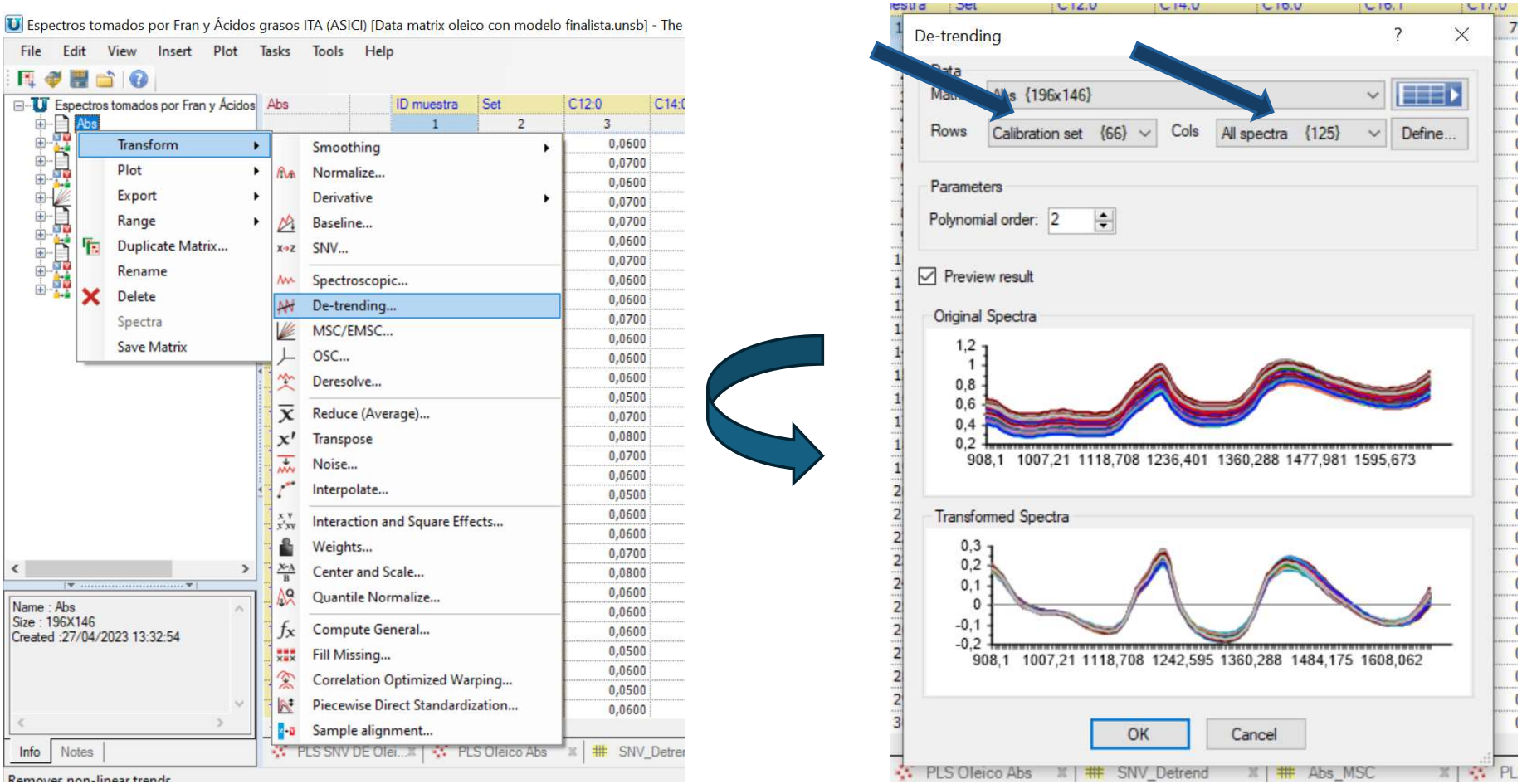
Detrending (DT)

Detrending (DT) is a transformation which seeks to remove nonlinear trends in spectroscopic data, normally caused by scatter effect.



- Pre-treatments

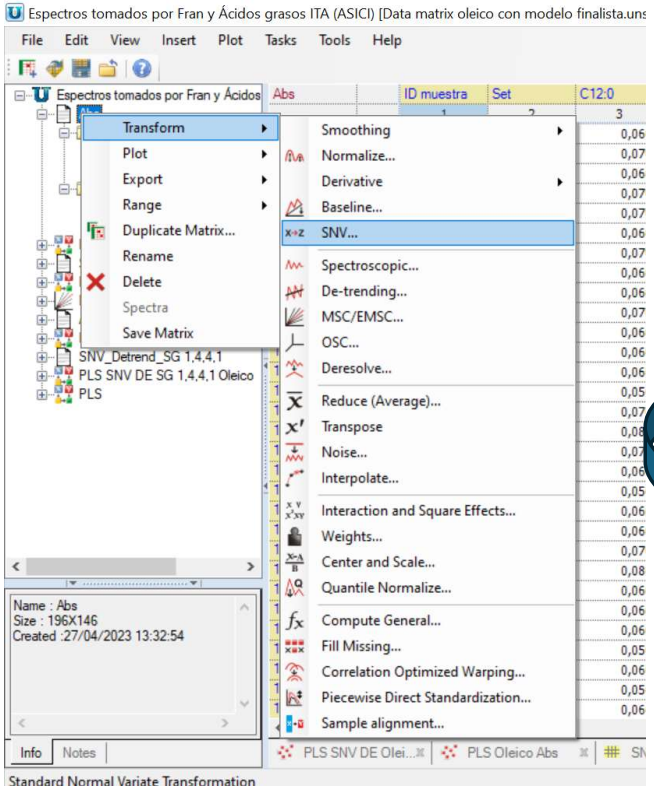
Detrending (DT)



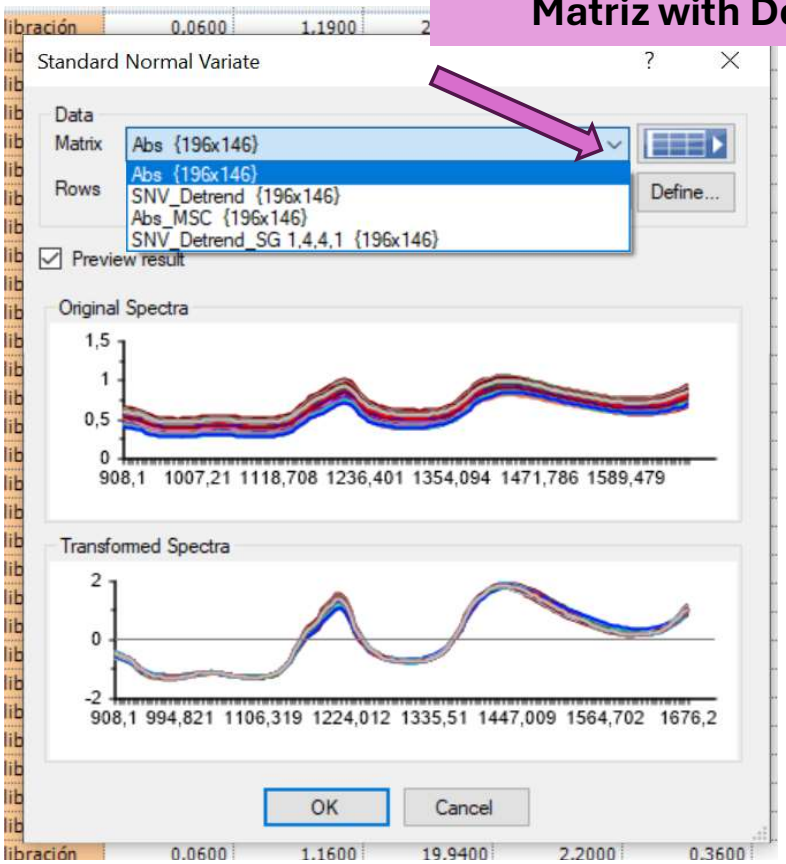
- Pre-treatments

Standard Normal Variate (SNV)

It is usually applied together with the DT. The SNV is mainly used to correct baseline variations caused by scattering.



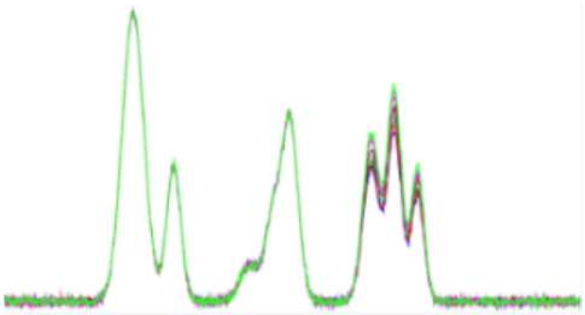
Matriz with Detrending pre-treatment



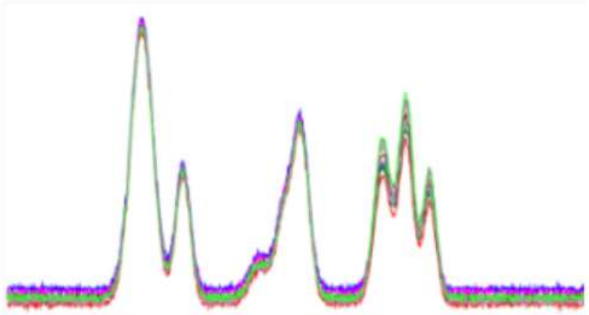
- Pre-treatments

Multiplicative Scatter Correction (MSC)

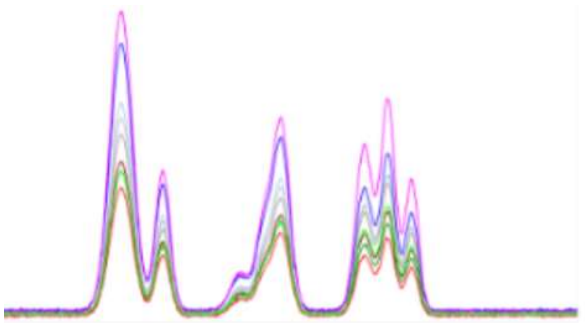
Multiplicative Scatter Correction (MSC) is a transformation method used **to compensate for pure additive and/or multiplicative effects** in spectral data.



Without scattering



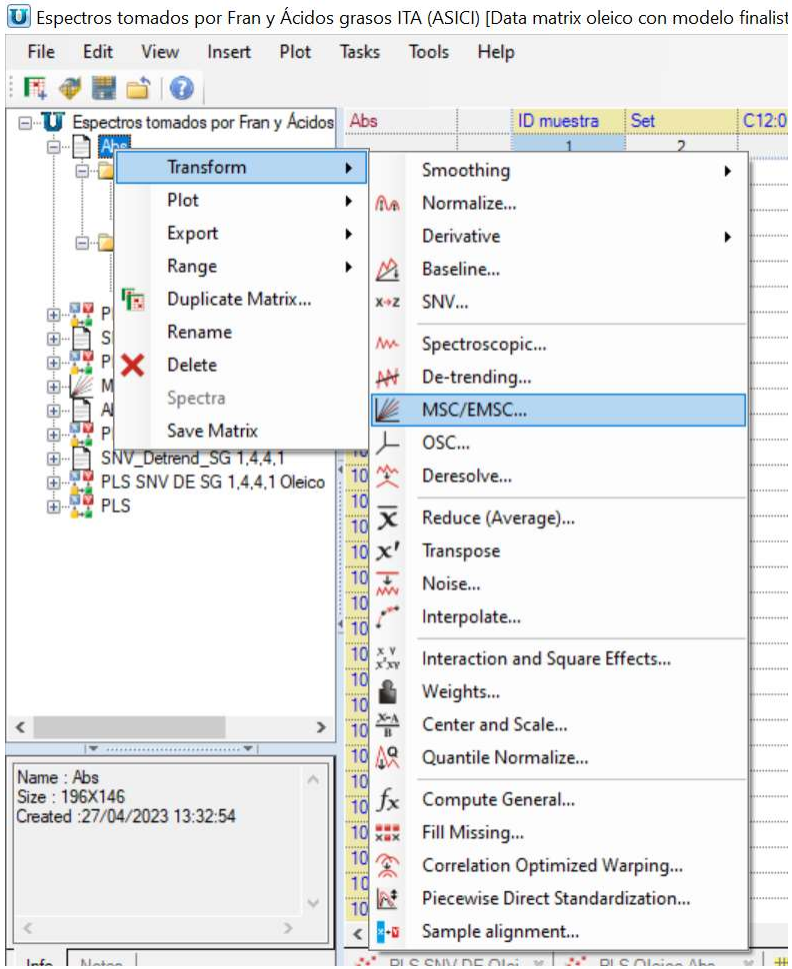
Additive baseline shift

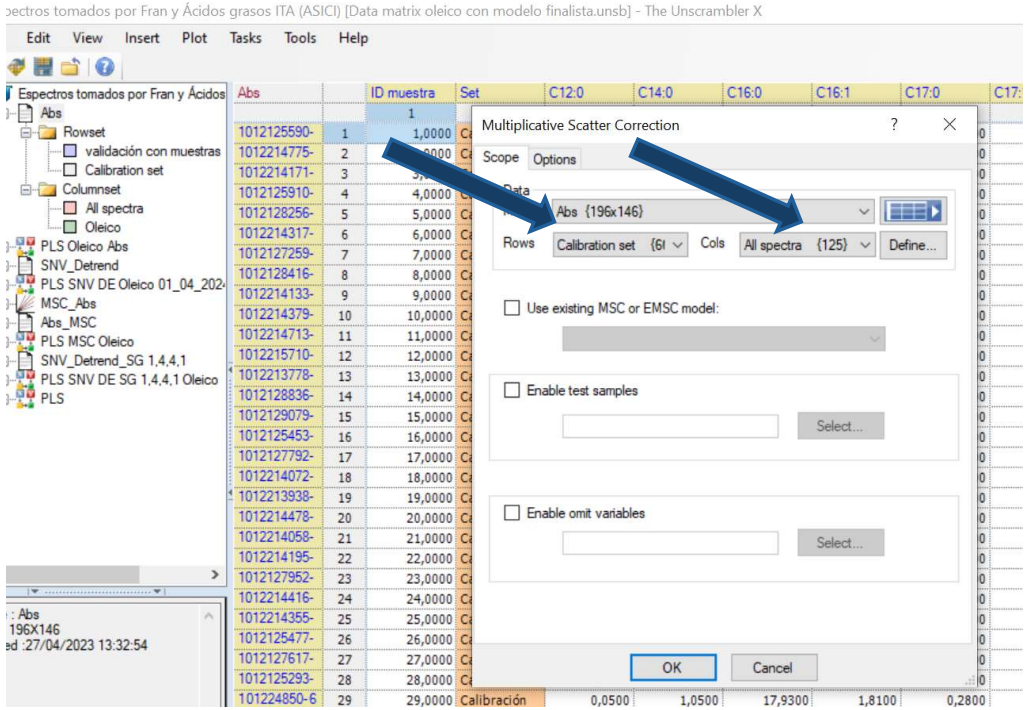


Additive and multiplicative effect

- Pre-treatments

Multiplicative Scatter Correction (MSC)



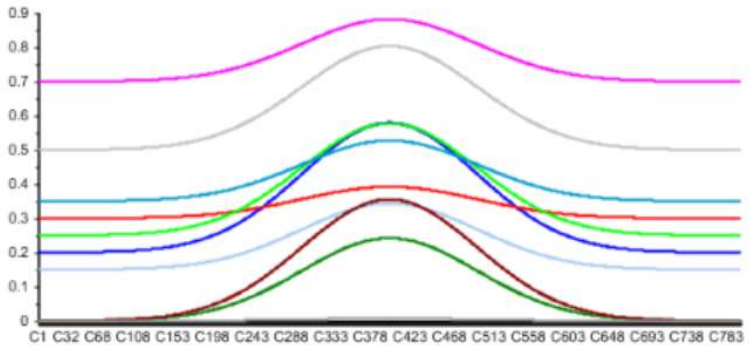


- Pre-treatments

Derivatives

Derivatives are applied **to correct for baseline effects (baseline shifts)** in spectra for the purpose of removing nonchemical effects. Derivatives may also aid in resolving overlapped bands which can provide a better understanding of the data, emphasizing small spectral variations not evident in the raw data.

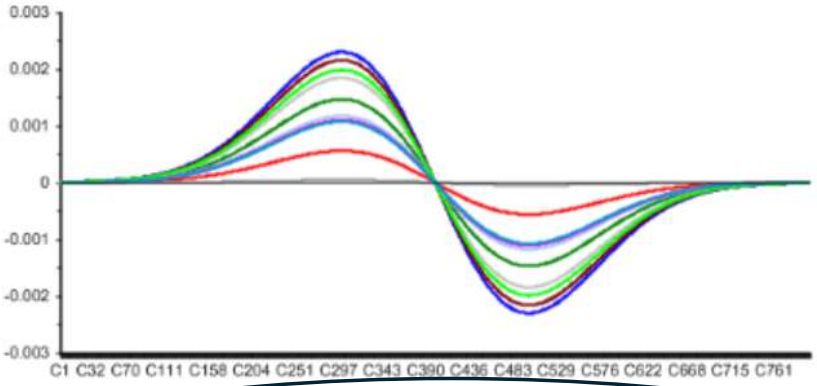
Raw data



The baseline offset has been removed under derivatization

The peak maxima in the raw data has now become a zero point in the derivative (The zero point can be explained by the fact that at a peak maxima (minima), the derivative is zero).

First derivate



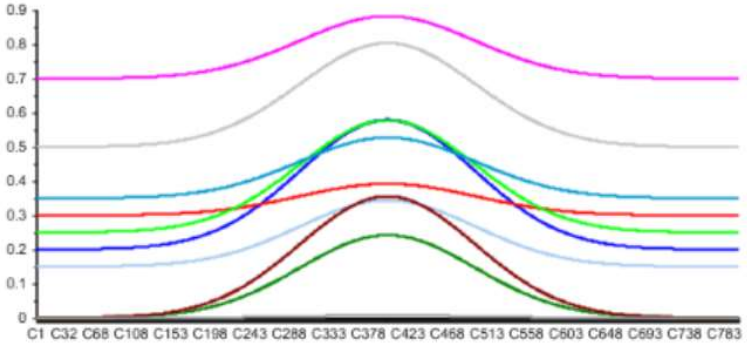
The **first derivative** eliminates **constant baseline shifts**

- Pre-treatments

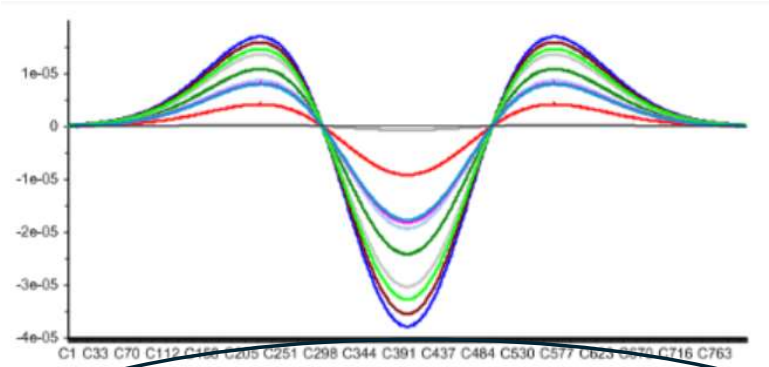
Derivatives

Derivatives are applied **to correct for baseline effects (baseline shifts)** in spectra for the purpose of removing nonchemical effects. Derivatives may also aid in resolving overlapped bands which can provide a better understanding of the data, emphasizing small spectral variations not evident in the raw data.

Raw data



Second derivate



The second derivative is a measure of the change in the slope of the curve. In addition to removing pure additive offset, it is not affected by any linear “tilt” that may exist in the data.

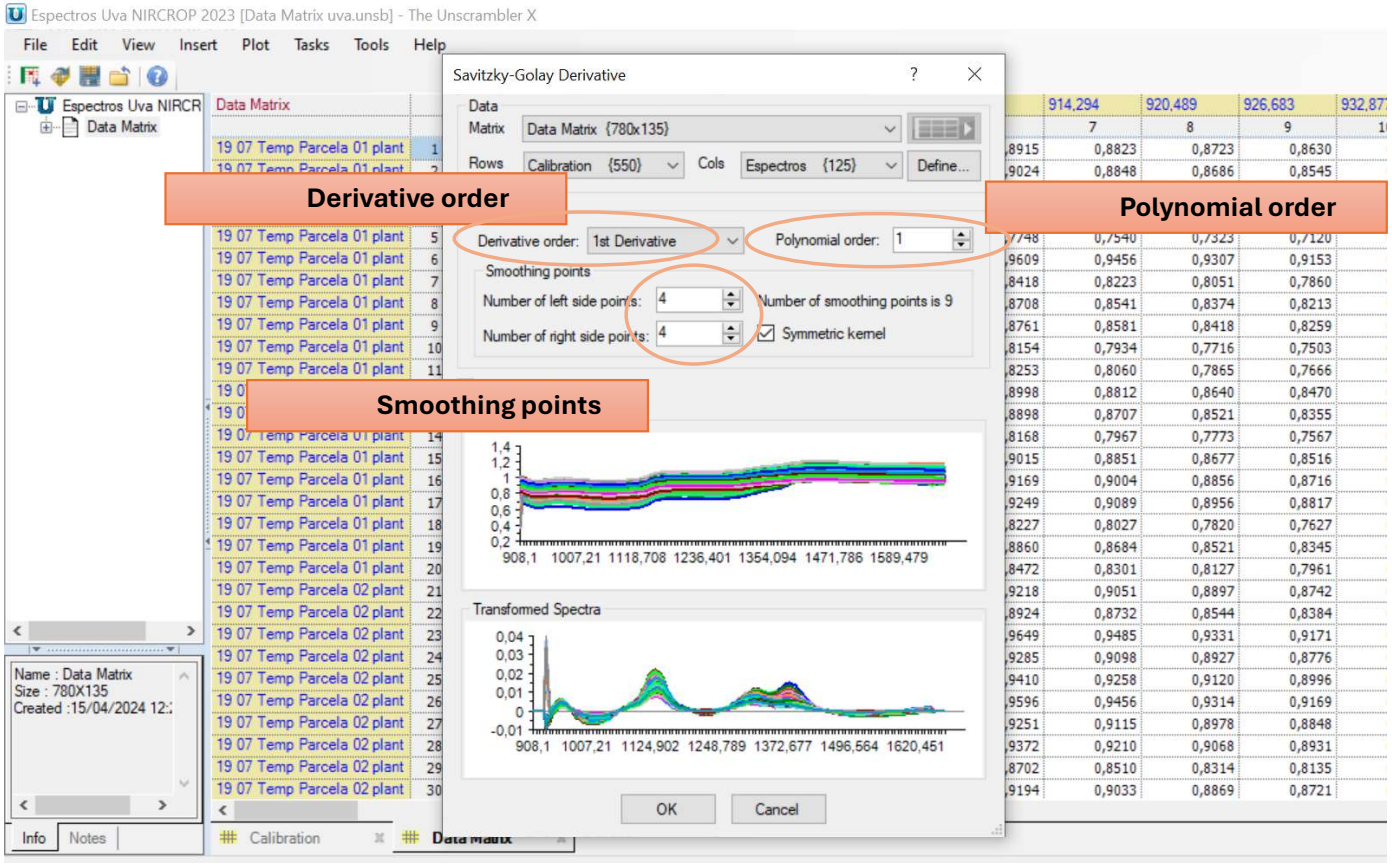
The second derivative can help resolve nearby peaks and **sharpen spectral features**

The **second derivative** eliminates **shifts that vary linearly with wavelength**

- Pre-treatments

Derivatives

The Savitzky-Golay method

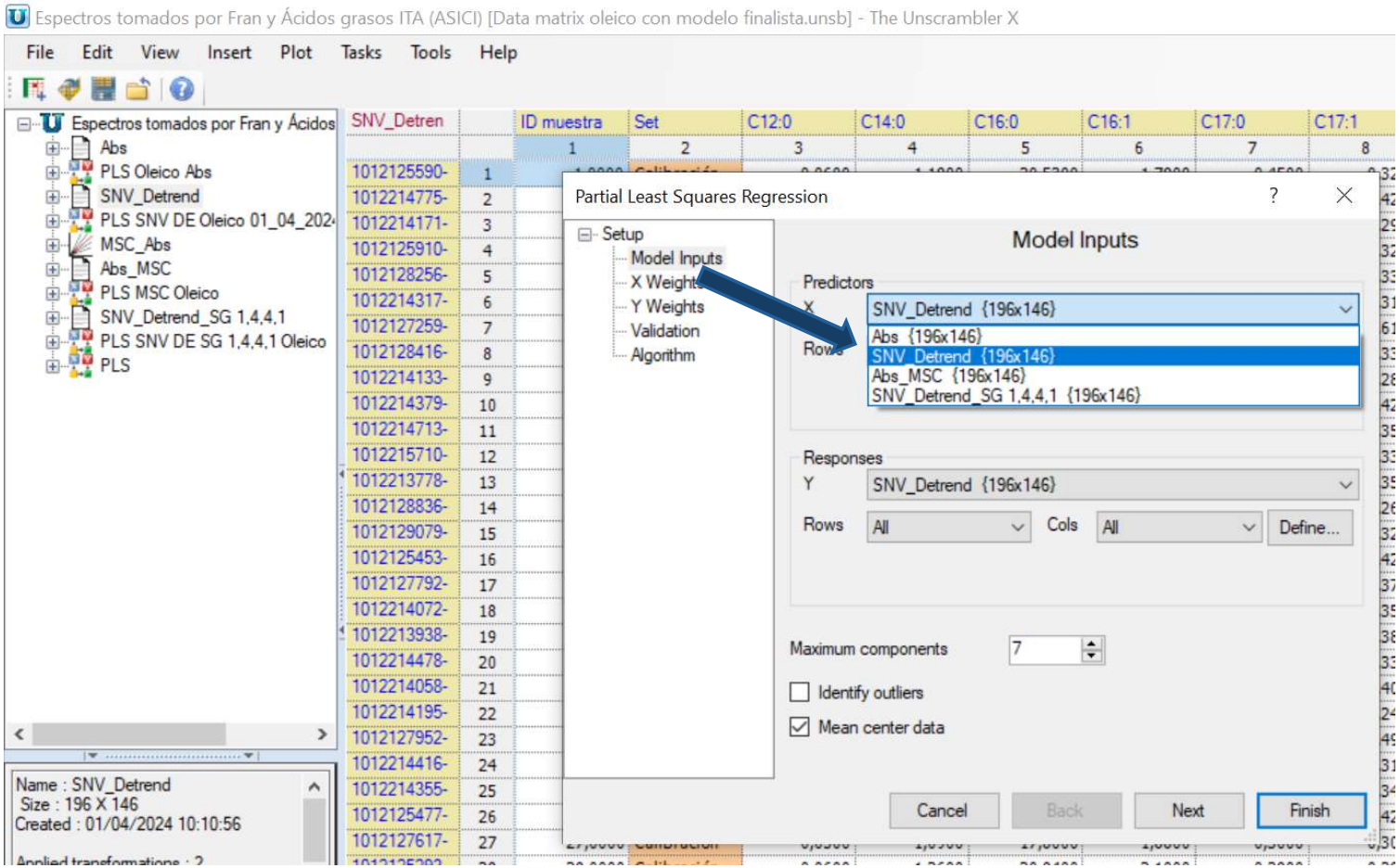


1,4,4,1

2,5,5,2

- Pre-treatments

Remake the model on the pretreated matrix



- Pre-treatments

Pretreatments are only applied to **spectral variables**

Pretreatments are applied to the **calibration set**

The selection of the most appropriate one is purely experimental

Archivos de Zootecnia
Journal website: https://www.uco.es/ucopress/az/index.php/az/

Comparación de herramientas quimiométricas de clasificación para la identificación de grasa perirrenal en corderos

Agudo, B.[1]; Delgado, J.V.[2]; López, M.M.[3] y Rodríguez, P.L.[4]

[1]Ovino del Suroeste Soc. Coop. Ltda. (OVISO), Villanueva de la Serena, Badajoz, España.
[2]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de Córdoba, España.
[3]Departamento de Producción Animal, CYCITEX, Finca La orden, Gobierno de Extremadura, España.
[4]Departamento de Producción Animal y Ciencia de los Alimentos, Facultad de Veterinaria de Cáceres, España.

Interreg

Cofinanciado por la Unión Europea

Cofinanciado pela União Europeia

España – Portugal

TID4AGRO

Desarrollo de modelos predictivos mediante tecnología NIRS

Tecnologías avanzadas, innovadoras y digitales para el sector agroalimentario de la EUROACE

Évora, 18th and 19th, novembre, 2024