BEYOND OPTIMIZE™ | 21

September 2021

Machine learning soft-sensors trained with Digital Twin for improving product quality & reducing energy

Daniel Salvador (CEPSA)
Alberto Pelaez (CEPSA)
JoseMaria Ferrer (Inprocess)

AGENDA

- Intro to CEPSA
- The plant revamp and the OTS
- Intro to Inprocess
- Generating datasets from HYSYS
- Creating the soft sensors
- Results and conclusions

About CEPSA

Cepsa is a **global, integrated** company operating across **the entire oil and gas value chain** and with over 90 years of experience.

Chemicals is one of the engines that is driving our internationalization and one of the areas where we are growing the most.

We are the **world leading producer** in **Alkylbenzene** (600 kMT/y total capacity)

Three LAB* Production sites located in San Roque (Spain), Becancour (Canada) and Camaçari (Brazil)

10,000 PROFESSIONALS

Who work with technical excellence, an ability to adapt, and an innovative spirit. They help us to stay competitive and are the key to face and overcome future challenges.

EXPERIENCE

We are supported by 90 years' experience in the oil world, making us a leading and robust company.

FIVE CONTINENTS

We operate across the entire oil value chain in five continents through our businesses in Exploration and Production, Chemicals, Refining, Distribution and Marketing, Gas & Power, and Trading.

INTERNATIONAL OUTLOOK

Our leadership in chemicals, combined with our broad experience in project execution across the world, helps us to grow internationally.

INTEGRATED BUSINESS

The physical integration of our production plants strengthens our model, reduces logistical costs and increases synergies.

SUSTAINABLE ENERGY

Our main priorities are to provide society with a safe, reliable and sustainable energy and contribute to the economic and social development of the communities where we work.

CUSTOMER SATISFACTION

Our commitment to the quality of our products and services and to the satisfaction of consumers forms the base of our customer relationship.

CHEMICALS LEADERSHIP

We are world leaders in LAB production (the raw material used to make biodegradable detergents), and in cumene. We are also the second world producer of phenol, used to make high performance materials, and acetone.

INNOVATION AND TECHNOLOGY

Innovation underpins everything we do. We have a Research Center that helps us to create value, optimize processes and improve the quality of operations and products.

- CEPSA is an integrated oil & gas Company and the largest Linear Alkylbenzene producer in the World (600 kMT/y)

*LAB- Linear Alkylbenzene

Puente Mayorga Petrochemical Plant produces LAB, raw material for biodegradable detergents. It supplies around 50% of the African and 25% of Western Europe markets

250 kt/y LAB

80 kt/y LABSA

400 kt/y Paraffins

100 kt/y Solvents

- Began operations in 1969 to produce LAB, raw material for Biodegradable Detergents
- Located at Algeciras Bay
- First plant ever to retrofit to Detal Technology (2021)
- Betting on quality: First company in Spain to achieve ISO 9001 certification (year 1992)

Source: Corporate Strategy

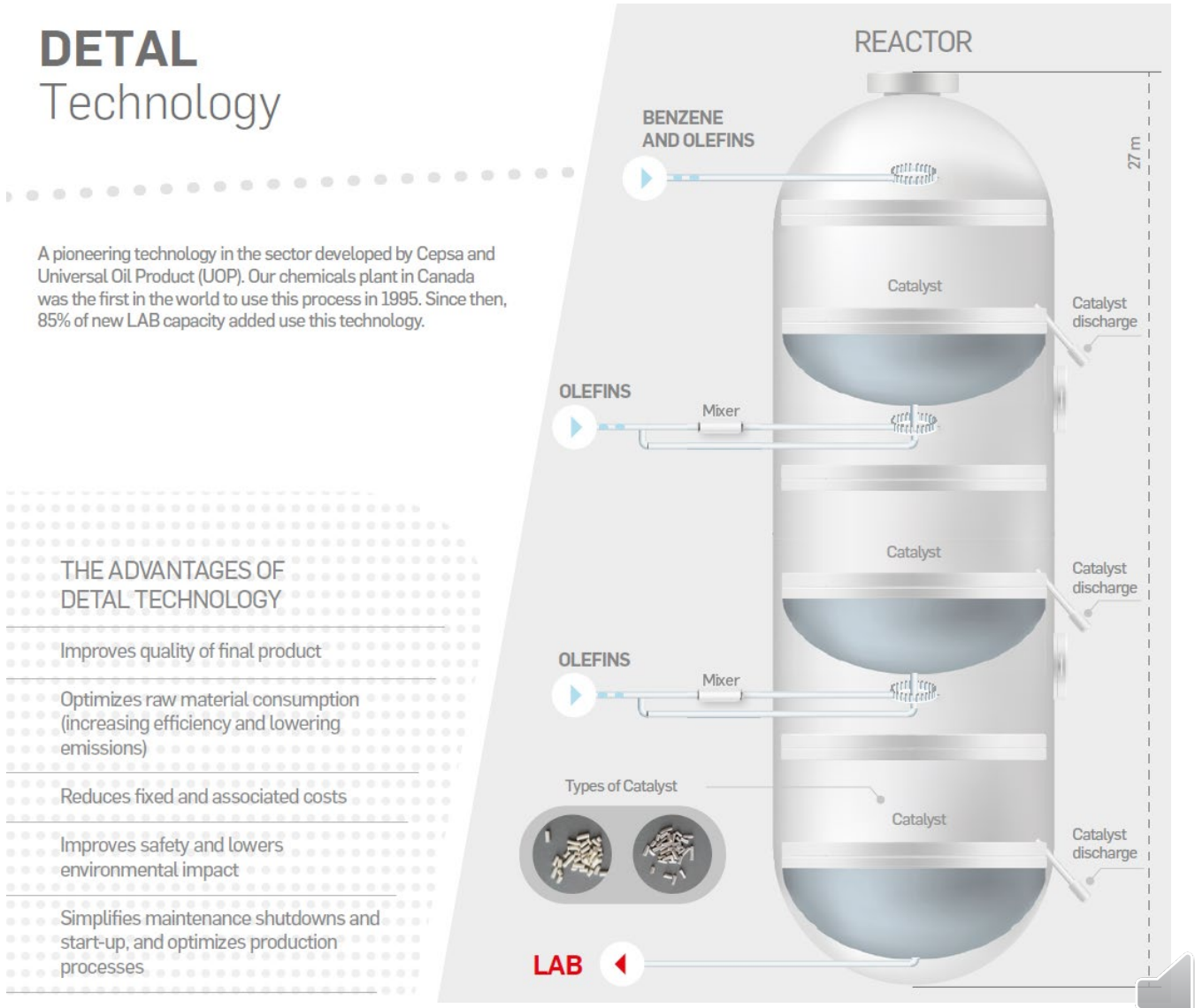
The capacity expansion and technology revamp

- Cepsa implemented a new upgrade project to expand production at its Puente Mayorga Chemical Plant in San Roque (Cadiz). The revamping process covers the installation of the new Detal technology, co-licensed by Cepsa and UOP, the most modern and efficient technology for the production of linear alkylbenzene (LAB), as well as increasing production capacity at the plant from 200 to 250 Kt/yr

https://www.youtube.com/watch?app=desktop&v=uDaLeUb0ij8

DETAL Technology

- The DETAL technology manufactures an improved product, enhance the efficiency of the plant and reduce emissions thanks to lower gas and power consumption and production process optimization.
- The new technology will also significantly improve safety since it is the first in the world that have changed the process from hydrofluoric hydric acid to DETAL technology.
- The new plant has two reactors to alternate production cycles and catalyst regeneration.



YET* 2.0: YET + DYNAMIC SIMULATION

LESSONS LEARNT:

Problems experienced on previously executed YET projects:

- ✓ Phenol 3 (CQP) – Fractionation train modeling: lack of data
- ✓ CQPM Paraffins project - C_{13} splitter modeling difficulties: lack of data
- ✓ The only possibility to increase the operating range of the RTO currently is by doing it before at the plant.

What if we do not need the plant running to create our dataset, but instead we create (simulate) them??

✓ Feasibility of predicting the quality of the separation in three specific columns using machine learning models trained with dynamic simulation data.

* Yield, Energy, Troughput

Inprocess in brief

independent from any ICSS or simulation software provider

our **core business** is Process Simulation

keen to **share its knowledge** with clients

2006

founded in Barcelona by domain experts

48 countries

worldwide presence

50+

simulation engineers

250+

years experience

400+

executed projects

320+

training courses

Mission: Deliver the value of Process Digital Twins to achieve Operations Excellence

Lifecycle Modelling and OTS projects

Simulation Lifecycle	Feasibility Studies	Dynamic Simulation Studies (DSS) Validation of Control	Operational Procedures Early OTS	DCS check-out	OTS Operator Training System	Plant start-up support	Operations & Maintenance support with the Digital Twin
Plant Lifecycle	Conceptual Design, Basic Engineering (FEED) and Detailed Engineering (EPC)					Commissioning & start-up	Operations & Maintenance

Lifecycle Modelling

- Feasibility studies. Selection of alternatives
- Dynamic simulation studies before plant construction
- Validation of control philosophy
- Operational procedures development/enhancement
- Process trainer - Emulated OTS - Early OTS
- DCS check-out
- OTS for operators' initial and continuous training
- Support during commissioning and start-up
- Operations & Maintenance support (Digital Twin)

Training & Knowledge Transfer

Training / Knowledge Transfer

- Process simulation courses
- Technology courses
- Knowledge Improvement Program – KIP
- Training for plant operators / technicians

Simulation Knowledge Profit
inprocess

Process Simulation projects

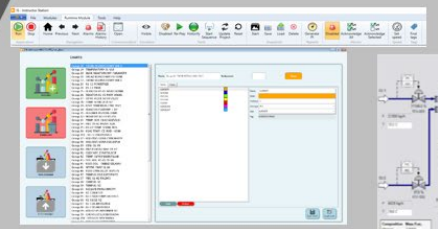
Process Simulation Studies

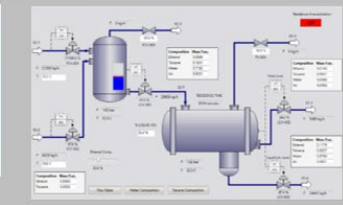
- Steady State Analysis
- Dynamic Simulation Studies
- Integrated Flare Systems Analysis
- Flow Assurance Studies with OLGA
- Utilities Network Models
- On-line models
- Operations Staff Training
- O&M Support

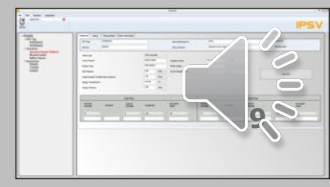
Applications & Software Development

Applications / Software Development

- **IIS**: Inprocess Infrastructure Suite
- **IPSV**: PSVs database
- **ITOP**: Inprocess Training for OPerators
- **ICOM**: Inprocess Competence Management System
- **IFLOW**: to link process simulators with OLGA®
- **IPSA**: Pressure Swing Adsorption simulator
- **OTS Web Access:** e-learning options
- Extensions for process simulators

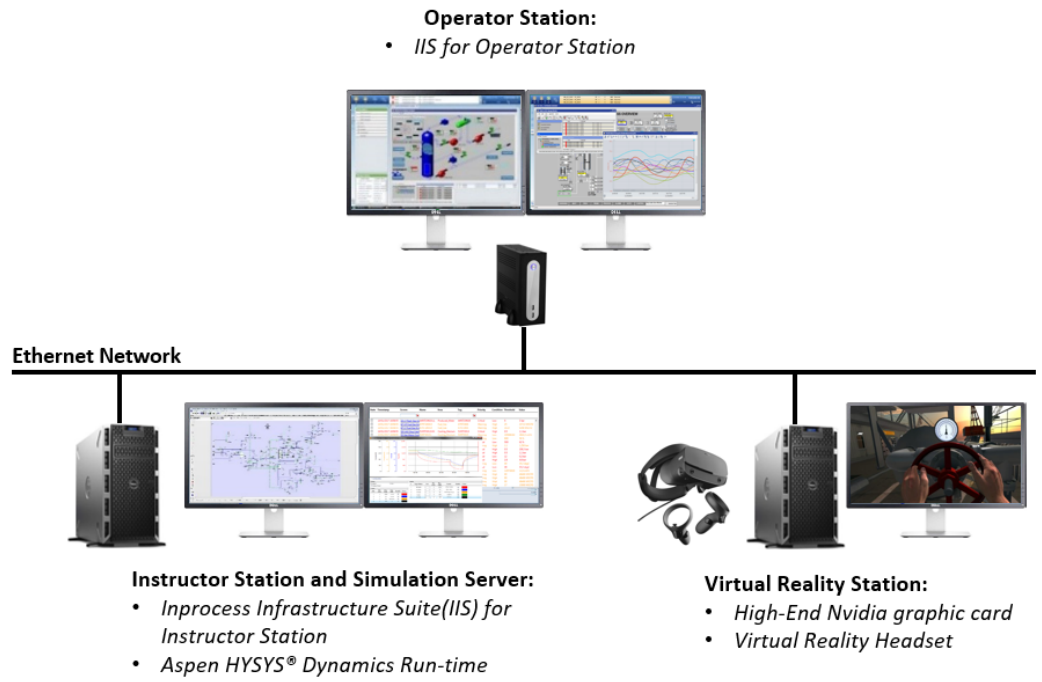


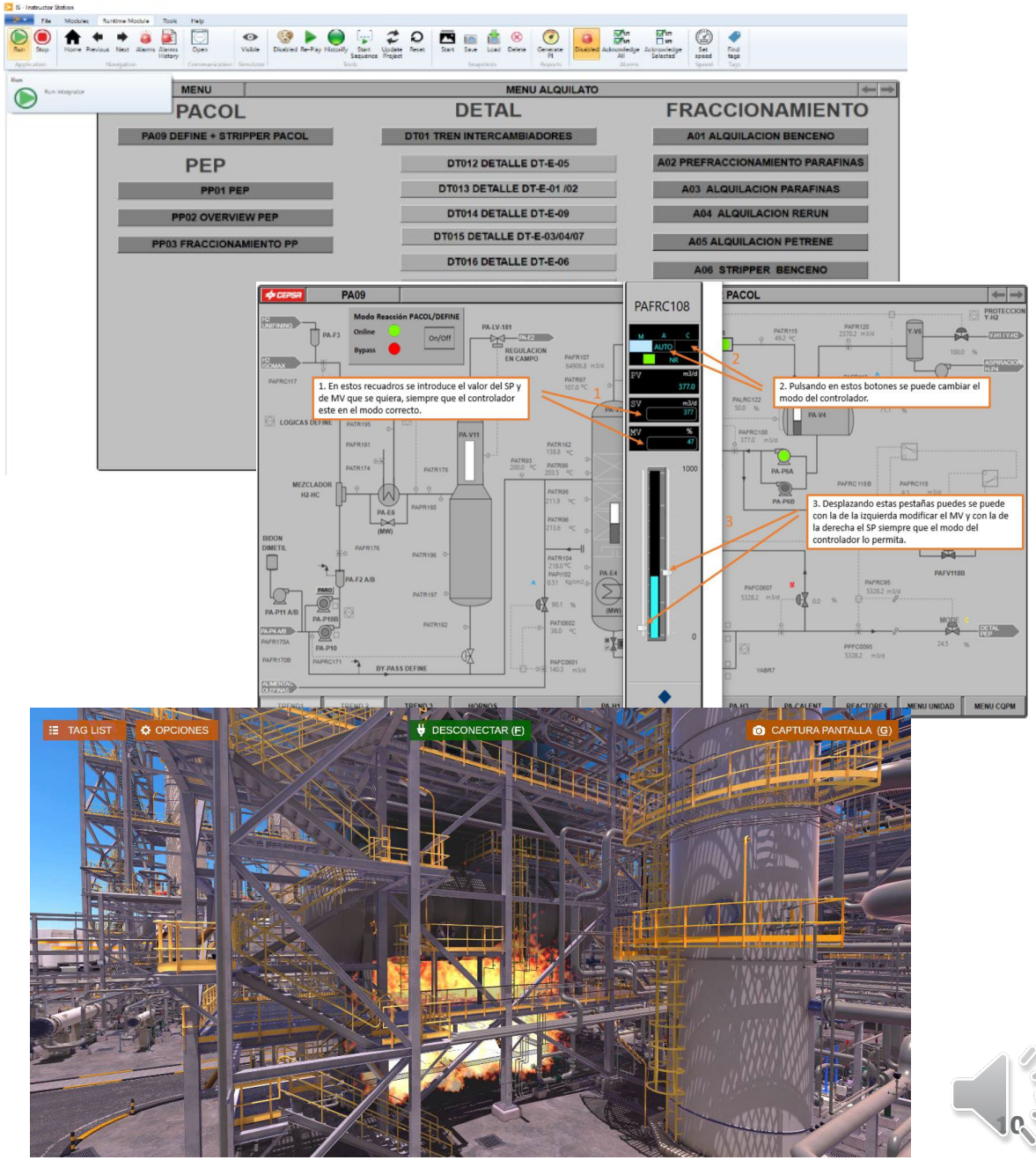




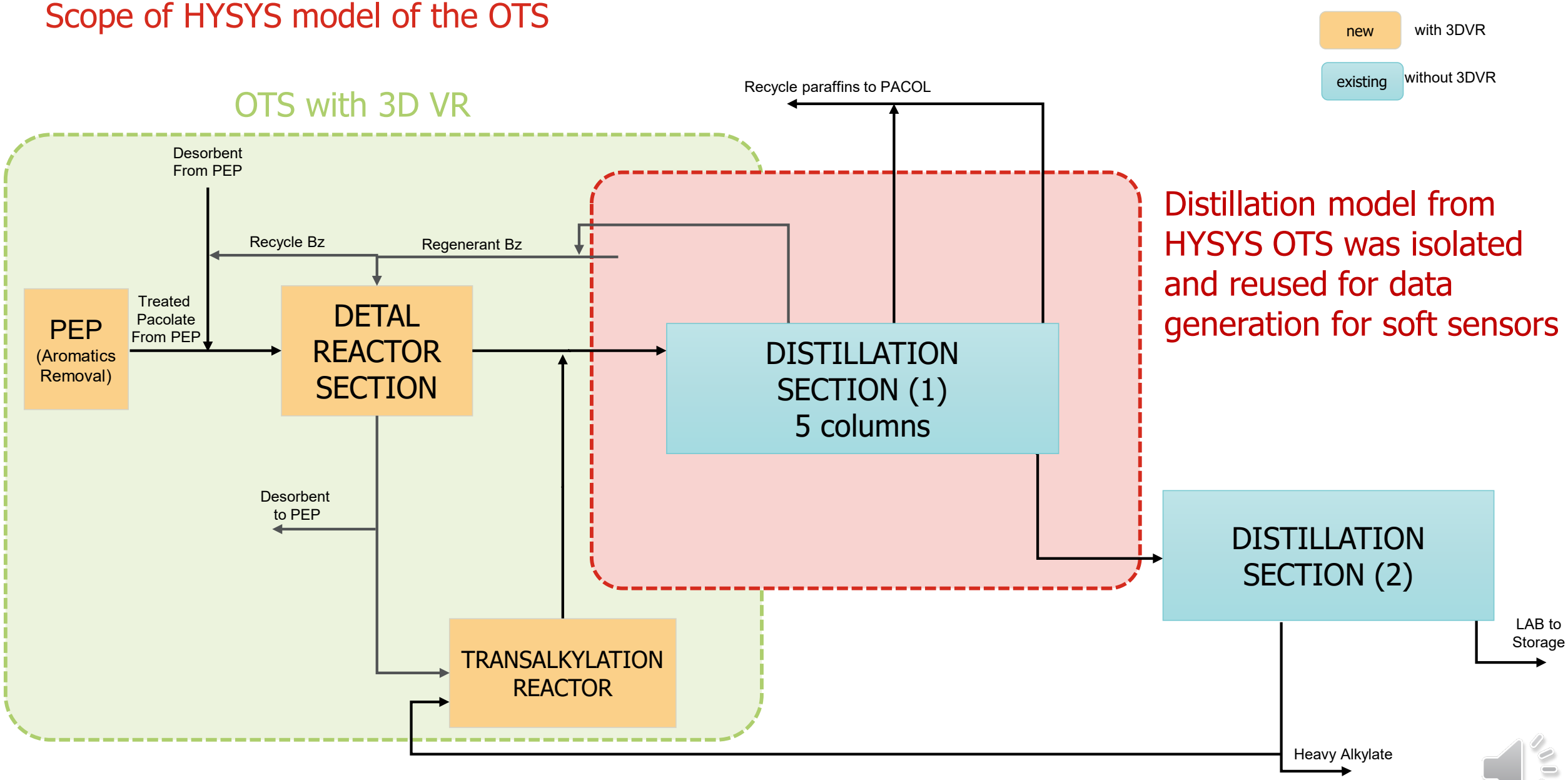
The OTS

- Inprocess was contracted to deliver an Operator Training Simulator (OTS) for Control Room and Field Operators with 3D Virtual Reality
  - Two different ICSS in one OTS
  - Complex 48 steps reactor regeneration sequences
  - Reuse of HYSYS models to develop ML models





Scope of HYSYS model of the OTS



Distillation model from HYSYS OTS was isolated and reused for data generation for soft sensors

Generation of data for soft sensors from HYSYS Dynamic models

Tested Feeds compositions table

	Case	Parafina (kmol/h)	LAB (kmol/h)	Benceno (kmol/h)
CARGA 10	1	809.66	117.05	1053.41
	2	809.66	117.05	1638.64
	3	800.66	128.04	1154.58
	4	800.66	128.04	1794.58
	5	791.67	139.04	1215.54
	6	791.67	139.04	1890.53
	7	782.67	144.03	1299.51
	8	782.67	144.03	2018.49
CARGA 11	9	880.83	128.04	1134.58
	10	880.83	128.04	1764.58
	11	880.73	139.84	1223.44
	12	880.73	139.84	1903.13
	13	870.83	145.83	1252.50
	14	870.83	145.83	1941.67
	15	860.94	151.73	1401.56
	16	860.94	151.73	2180.21
CARGA 12	17	871.98	139.04	1215.54
	18	871.98	139.04	1890.53
	19	860.80	145.83	1252.50
	20	860.80	145.83	1941.67
	21	850.00	256.83	1409.66
	22	850.00	256.83	2192.80
	23	838.20	187.42	1606.82
	24	838.20	187.42	2543.84

Tested Input variables tables

DT-V-03

Zona	Variables	Tag	Valor Sacyr	Valor Train1	Valor Train2	Valor Train3	unid.	Rango Min	Rango Max	Salto
Carga	Carga PACOL (1=475 m3/d) (corriente 265)							10	12	1
	Ole/Para (mol:mol) (co							10	13	1
	Ra									
	Ma									
Cabeza	Tem									
Energía										

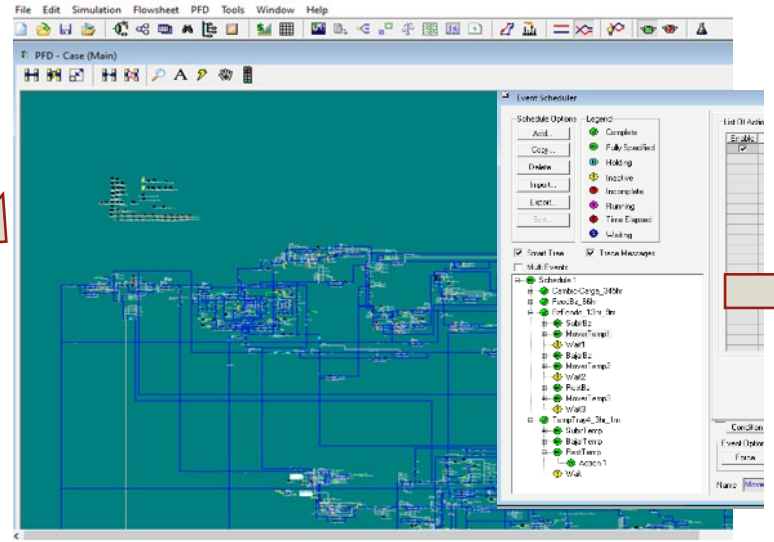
ColBz (AV-11/A-V110)

Zona	Variables	Tag	Valor Sacyr	unidades	Rango Min	Rango Max	Salto
Carga	6 casos de carga: variación LAB/Paraf						
Cabeza							
Energía	Bzfo						
Presion							

AV-51/AV14

Zona	Variables	Tag	Valor Sacyr	unidades	Rango Min	Rango Max	Salto
Carga	3 Casos carga						
Cabeza	Reflujo A-V51	AFRC501.SP	175.4	m3/d			10
	Reflujo A-V14	AFRC137.SP	1063	m3/d			10
	Presión A-V14 (mmHg)	APRC138.SP	750		440	750	10 var. 10 down
Energía	Temp plato24 --> Rebcolum.paraf. A-V51	ATRC504.SP	132.4	Deg C			1
	Temp plato34 --> Rebcolum.paraf. A-V36	ATRC548.SP	235	Deg C			2

HYSYS Dynamic model of the four columns studied



Models were split in several cores running for 3-4 days at 15 Real time factor to represent about 1600 different plant status.
CEPSA was also taught to run their own runs.

Calculated Exported variables

Columna Prefraccionadora de Benceno (DT-V-03)

Zona	Variable	Tag	Descripción
Carga	Caudal	Caudal corriente 265	Carga a DT-V-3
		Caudal corriente 753	LAB de Transalk
		Caudal corriente 340	Efluente regenerador a DT-V03
	Temperatura	DTTI773	Carga a DT-V-3
		DTTI861	LAB de Transalk
		DTTI783	Efluente regenerador a DT-V03
Reflujo	Caudal	DTFC821	Reflujo caliente
	Temperatur		
	Nivel		
Cabeza	Temperatur		
	Nivel		
	Presión		
Fondo	Caudal		
	Temperatur		
	Nivel		
	Presión		
Energía	Caudal		
	Temperatur		
Columna	Temperatur		
	Presión		
Caudales molares y composiciones	Composic.		
	Caudal mol a		
	Caudal molar c		
	Composic.		

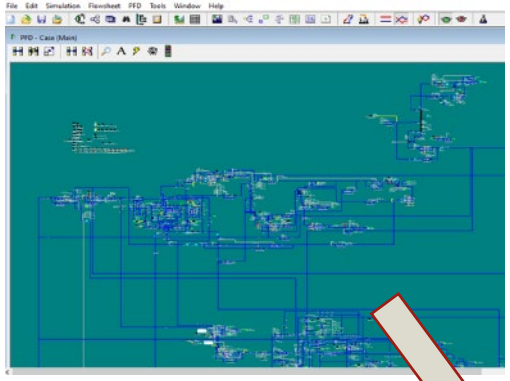
Stripper Parafinas (A-V51)

Zona	Variable	Tag	Descripción
Reflujo	Caudal	AFRC501	Caudal reflujo
		AFRC504	Caudal hacia aspiración A-P10A/B
	Temperatura	ATRC160	Temperatura reflujo
		ATR512	Temperatura A-V53
		ATR511	Temperatura salida carcasa A-E53
		ATR513	Temperatura hacia aspiración A-P10A/B
	Nivel	ALRC502	Nivel A-V53
	Presión	APR503	Presión A-V53
Cabeza	Temperatura	ATR526	Temperatura cabeza
	Presión	APRC501	Presión cabeza
Fondo	Caudal	AFRC503	Caudal fondo
	Nivel	ALRC501	Nivel fondo
Energía	Caudal	AFRC502	Caudal tubos A-E51
	Temperatura	ATR507	Temperatura salida tubos A-E51
Columna	Temp		
	Pr		
Composic.			

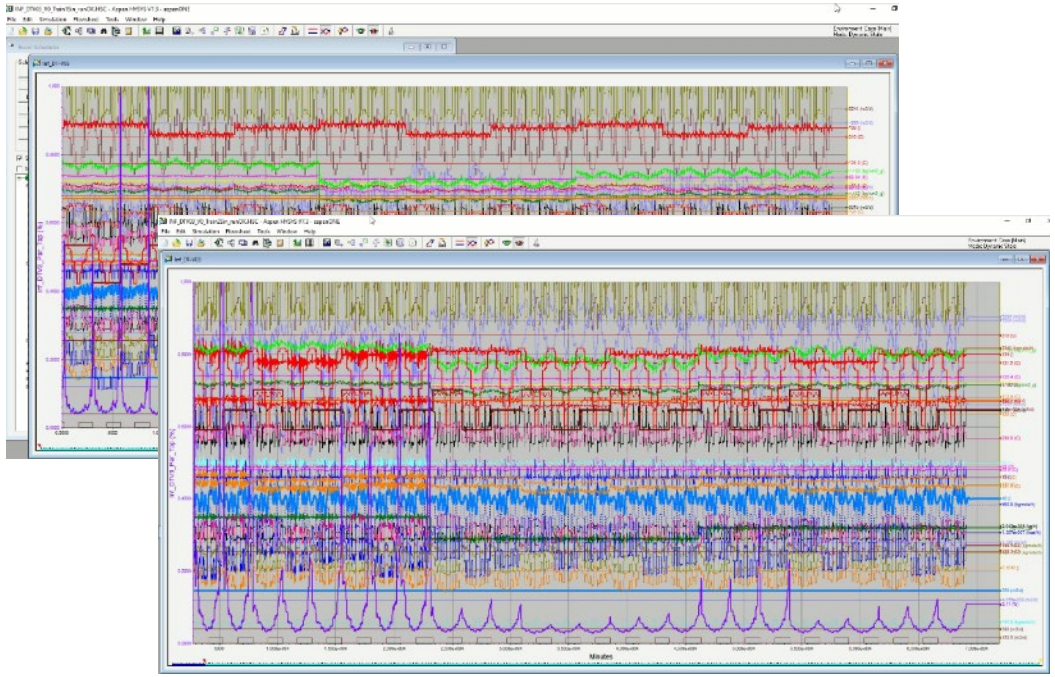
Columna Benceno (A-V11 & A-V110)

	Zona	Variable	Tag	Descripción
A-V11	Carga	Caudal	AFRC46	Parafinas de bombas de anillo líquido
		Temperatura	ATR334	TVapores A-V110 a A-V11
	Reflujo	Caudal	AFC131	Caudal de reflujo
			AFC26	Caudal hacia succión DT-P-04 A/B
		Temperatura	ATR134	Temperatura de reflujo
		Nivel	ALRC137	Nivel botellon de reflujo
	Cabeza	Temperatura	ATR125	Temperatura cabeza
		Presión	API834 y APR11	Presión cabeza
	Fondo	Caudal	DTFI341	Caudal fondo
		Temperatura	ATI870	Temperatura fondo
			ATRC509	Temperatura hacia A-V110
		Nivel	ALRC118	Presión en el nivel de fondo
	Columna	Temperatura	ATR123	Temperaturas columna
			ATRC124	
			ATR120	
A-V110	Fondo	Caudal	AFRC140	Caudal fondo
		Temperatura	ATR229	Temperatura fondo
		Nivel	ALRC330	Nivel fondo
	Energía	Caudal	AFRC122	Caudal tubos A-E15
			AFRC327	Caudal tubos A-E150
		Temperatura	ATR121	Temperatura salida tubos A-E15
			ATR329	Temperatura salida tubos A-E150
			ATRC509	Temperatura salida carcasa A-E52
			ATR508	Temperatura salida tubos A-E52
	Columna	Temperatura	ATR333	Temperaturas columna
			ATR332	
			APR110	Presión cabeza
Composiciones y caudales modelares	Composic.		INF_AV11_PAR_TOP	Total Parafinas en cabeza columna A-V11 (p
			INF_AV110_BZ_BOT	Benceno en fondo columna A-V110(ppm
	Caudal Molar		462_Parafinas	Caudal de parafinas en corriente 462 (fond
			462_LAB	Caudal de LAB en corriente 462 (fondo)

Virtual process data generated by HYSYS Dynamics tests

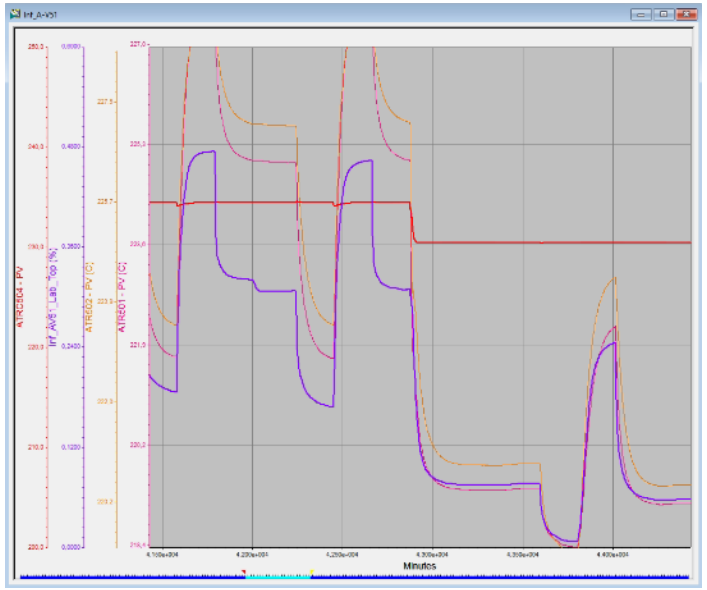


The data contained key compositions (impurities in top or bottom of distillation columns) which are not directly measured in the plant



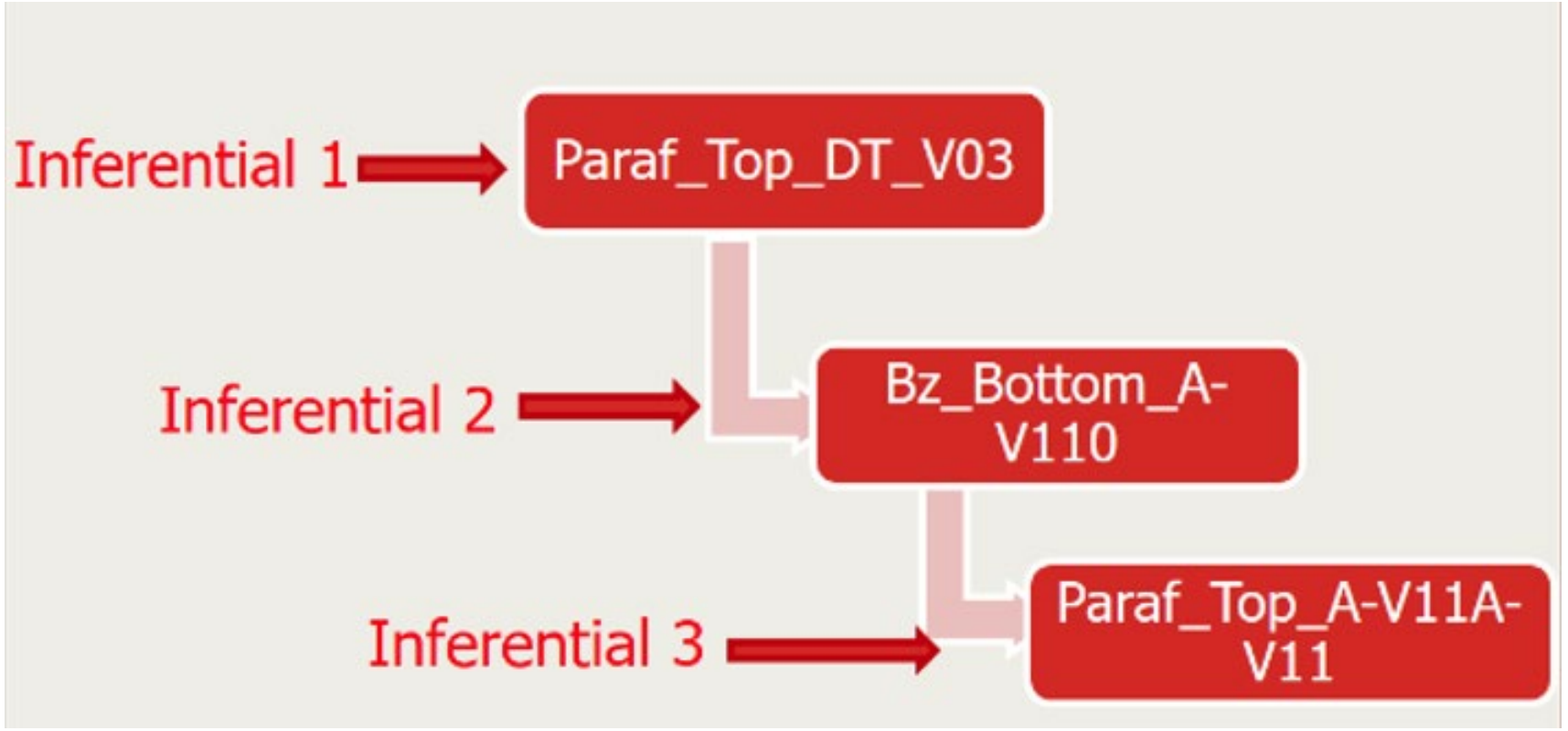
Exported Data

- INF_DTV03_Train1.csv
- INF_DTV03_Train2.csv
- Inf_ColBz.csv
- Inf_ColBz_PHigh0.7.csv
- Inf_AV14_PLow-710.csv
- Inf_AV51_PLow-710.csv
- Inf_AV14_Phigh-680.csv
- Inf_AV51_Phigh-680.csv
- Inf_AV14_CARGA12_caso17.csv
- Inf_AV51_CARGA12_caso17.csv
- Inf_AV14_CARGA10_caso7.csv
- Inf_AV51_CARGA10_caso7.csv



CEPSA used their Data Scientists to train their Machine Learning models and deploy them in the plant

Machine learning models



Machine Learning modeling methodology

- Machine Learning is a branch of the Artificial Intelligence whose goal is to build systems that automatically learn from data and adjust actions or decission accordingly.

Cross-industry standard process for data mining (CRISP-DM)

Software

Python was used to do the exploratory analysis, data preparation and models of this project.

Advantages:

- Open source.
- Easy to use.
- Flexible and fast to perform flexible exploratory analysis.
- Easy to find internet errors helps

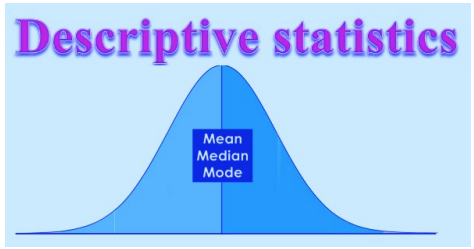
Disadvantages:

- The need of programing skills

Note: Aspen Hybrid Models were not available at the time of the project.

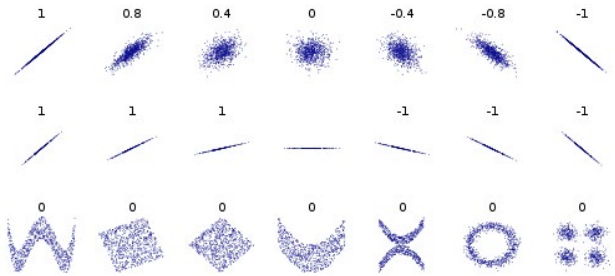
Data discovery and analytics

Descriptive Statistic



Summarizing and organizing the data so they can be easily understood

Correlated variables



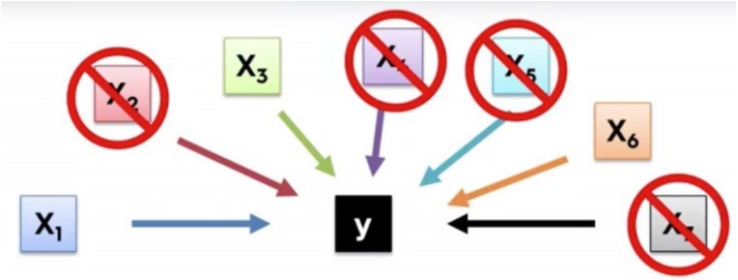
The workflow to determine variables correlation was:
1- Set group of correlated variables
2- Select the most important variable with respect to the objective and eliminate the rest of them

Variables importance

The feature selection is the process that choose a reduced number of explanatory variable to describe a response variable.
1- Make the model easy to interpret
2- Reduce the size of the problem allowing algorithms to work faster
3- Reduce overfiting

- ✓ *Lasso*
- ✓ *Gradient Boosting*
- ✓ *Random Forest*
- ✓ *Recursive Hybrid elimination**

* Several times iterated



Variables group

The groups were built based on the importance of the variables and the knowlege of the business

Data preparation

Bz_Bottom_A-V110:

- 33 Variables (columns) generated by Hysys
- 85,848 rows
- All numerical variables
- Determination of minimum, maximum, average and standard deviation values, distribution histograms.

Correlated variables:

9 groups of correlated variables were determined.

Determination of the most important variables:

- ✓ *Lasso*
- ✓ *Gradient Boosting*
- ✓ *Random Forest*
- ✓ *Recursive Hybrid elimination**

* Several times iterated

Based on the observed before and knowledge of business, the following groups of variables were determined:

Group 1
API0834
ATR333
ATRC332
ATR134

***Group 2:**

Group 1 – ATRC333**

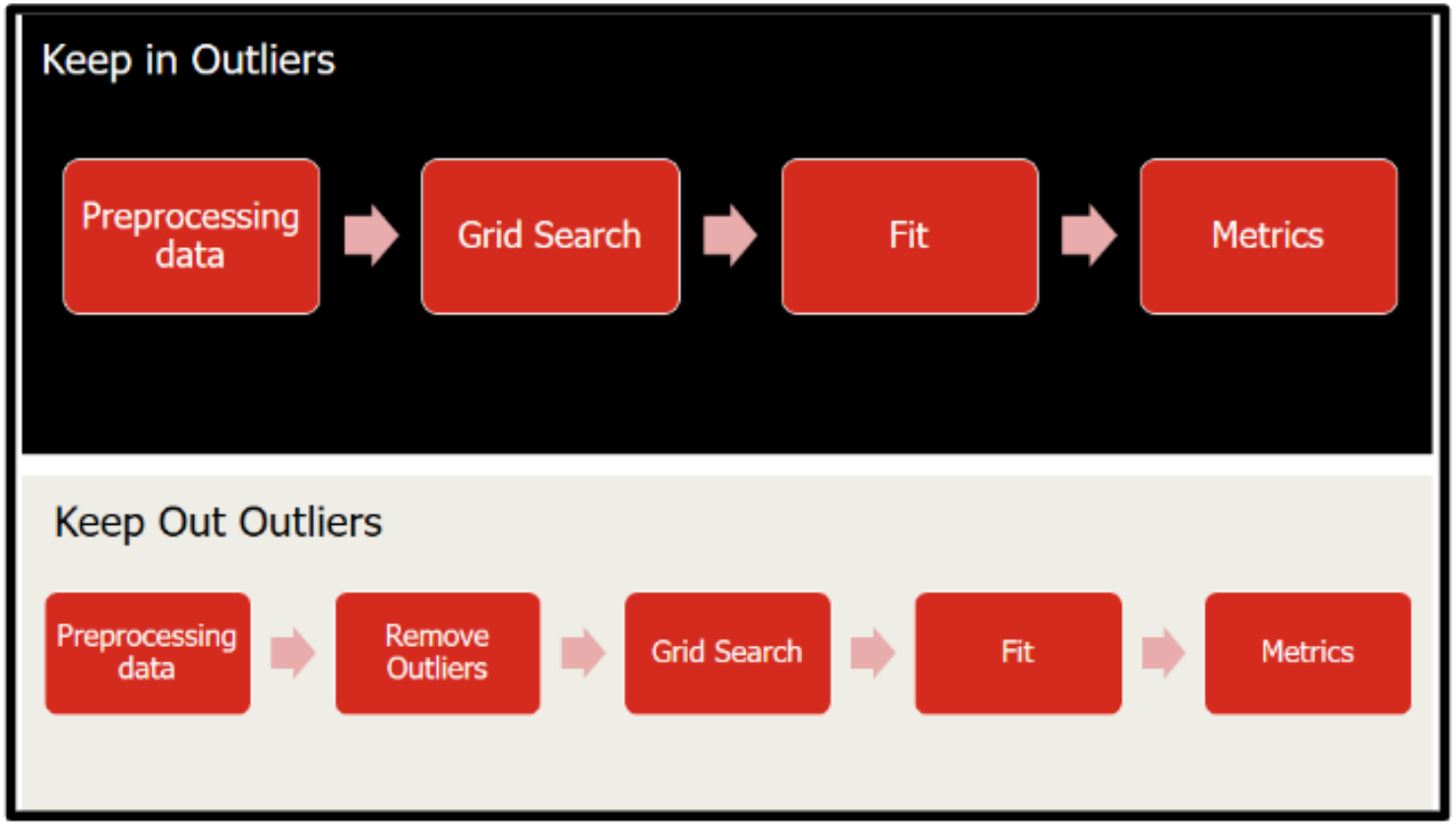
**Sensor Wifi

Group 3
API0834-ATR134/ATRC332-ATR134
ATRC332
ATRC332-ATR134/API0834

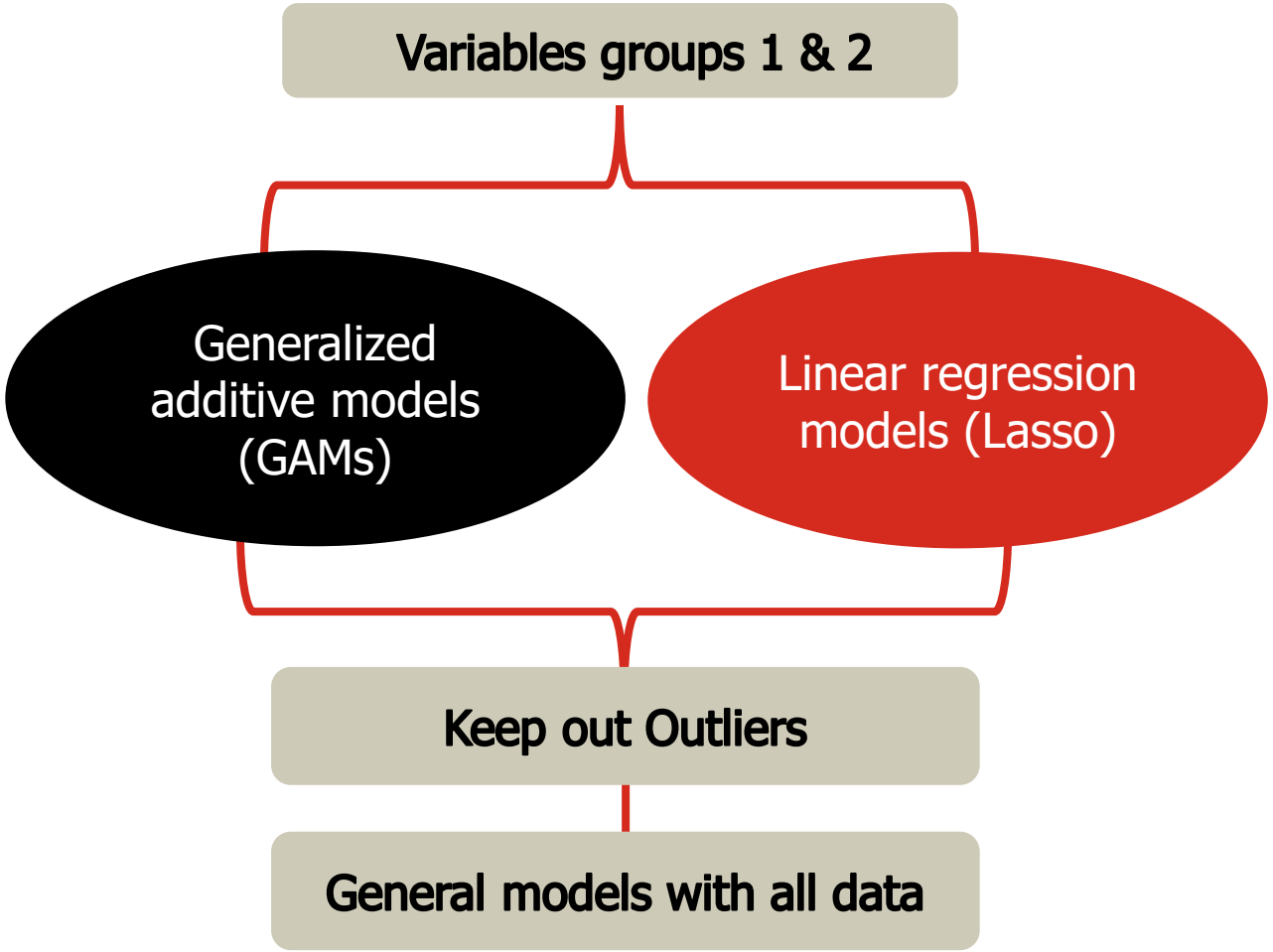
Group 4
API0834
API0834-ATR134/ATRC332-ATR134
ATRC332
ATR134

The groups were built based on the correlations of the variables with the target and process knowledge.

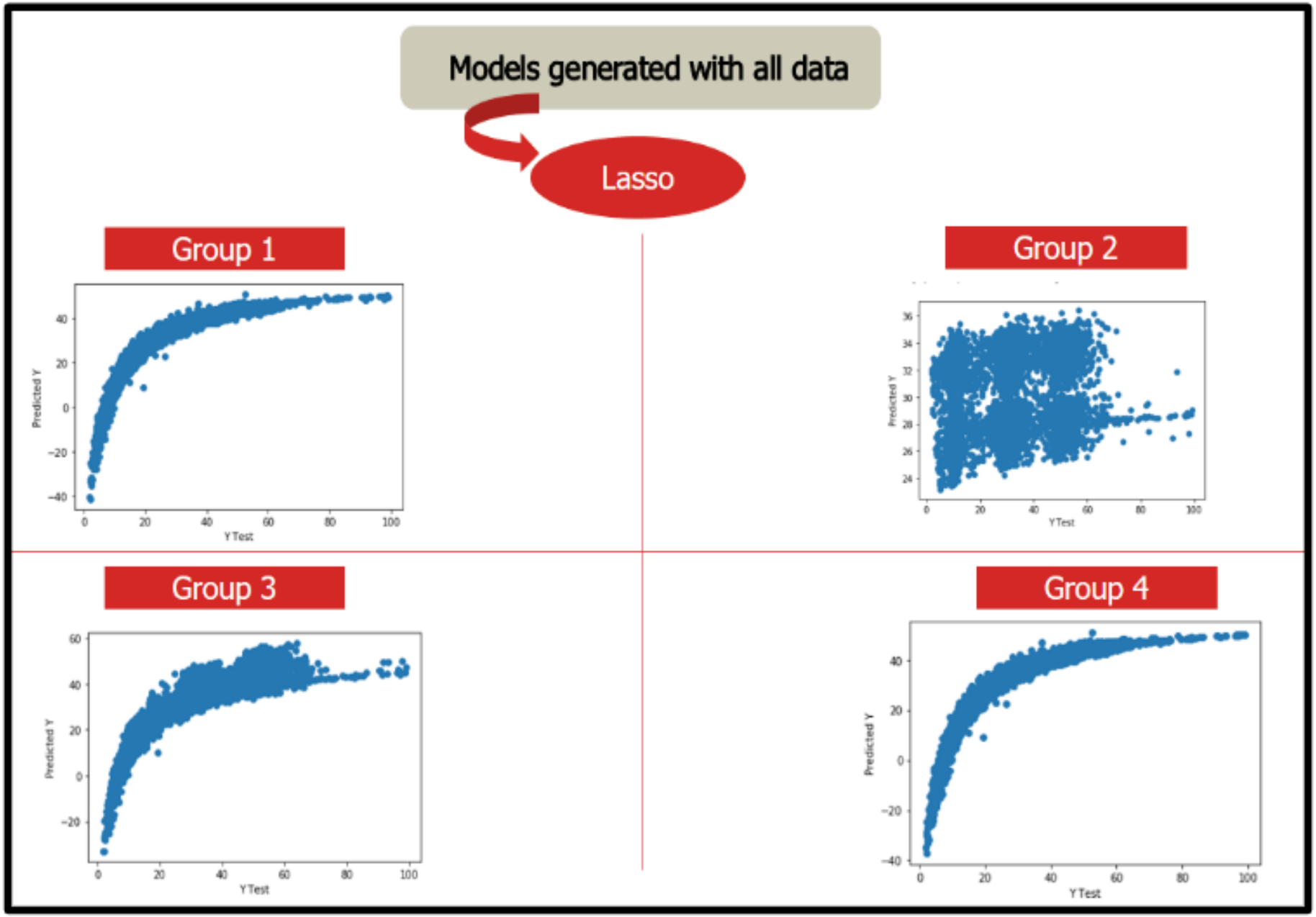
Machine learning modeling



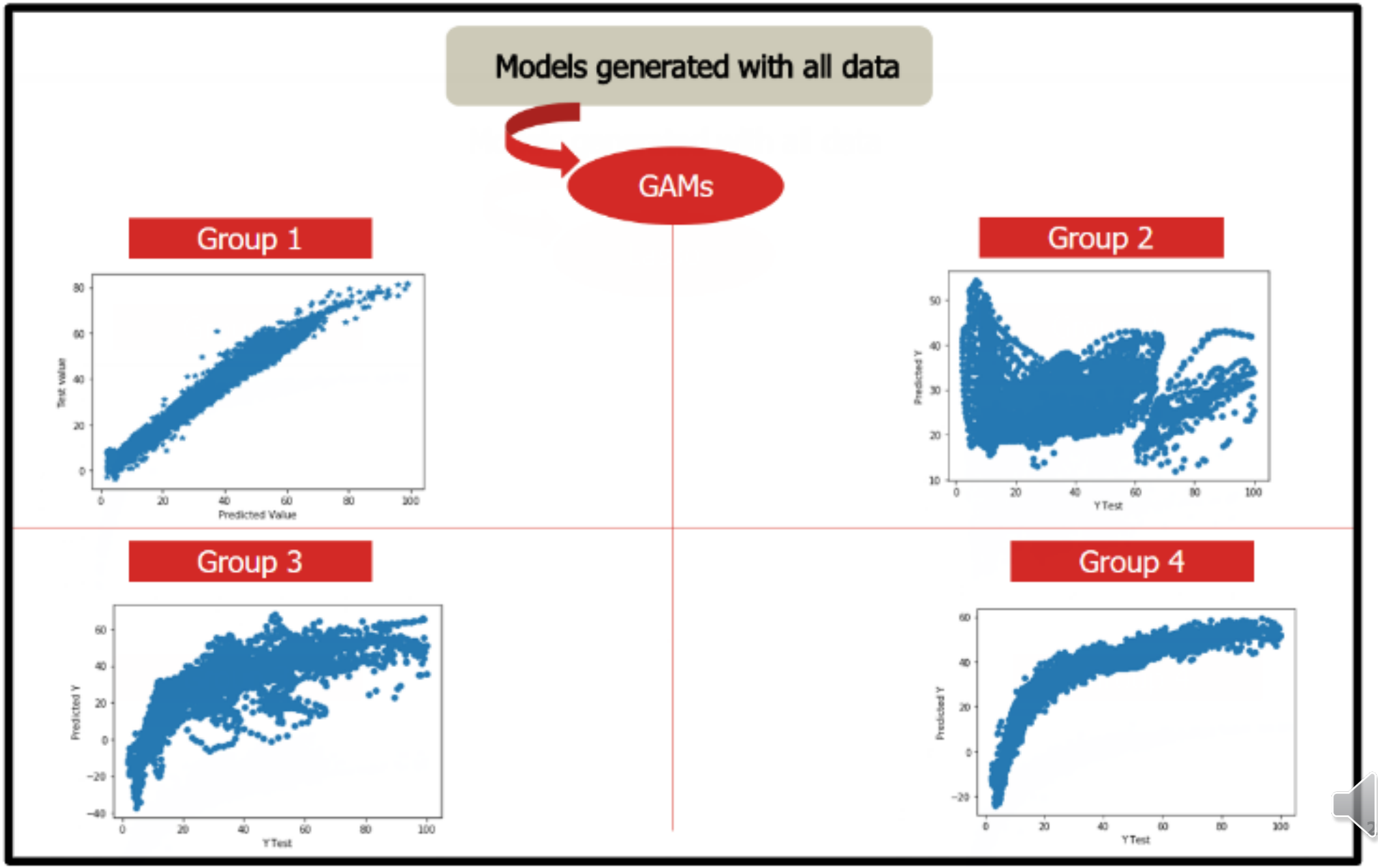
Machine learning modeling



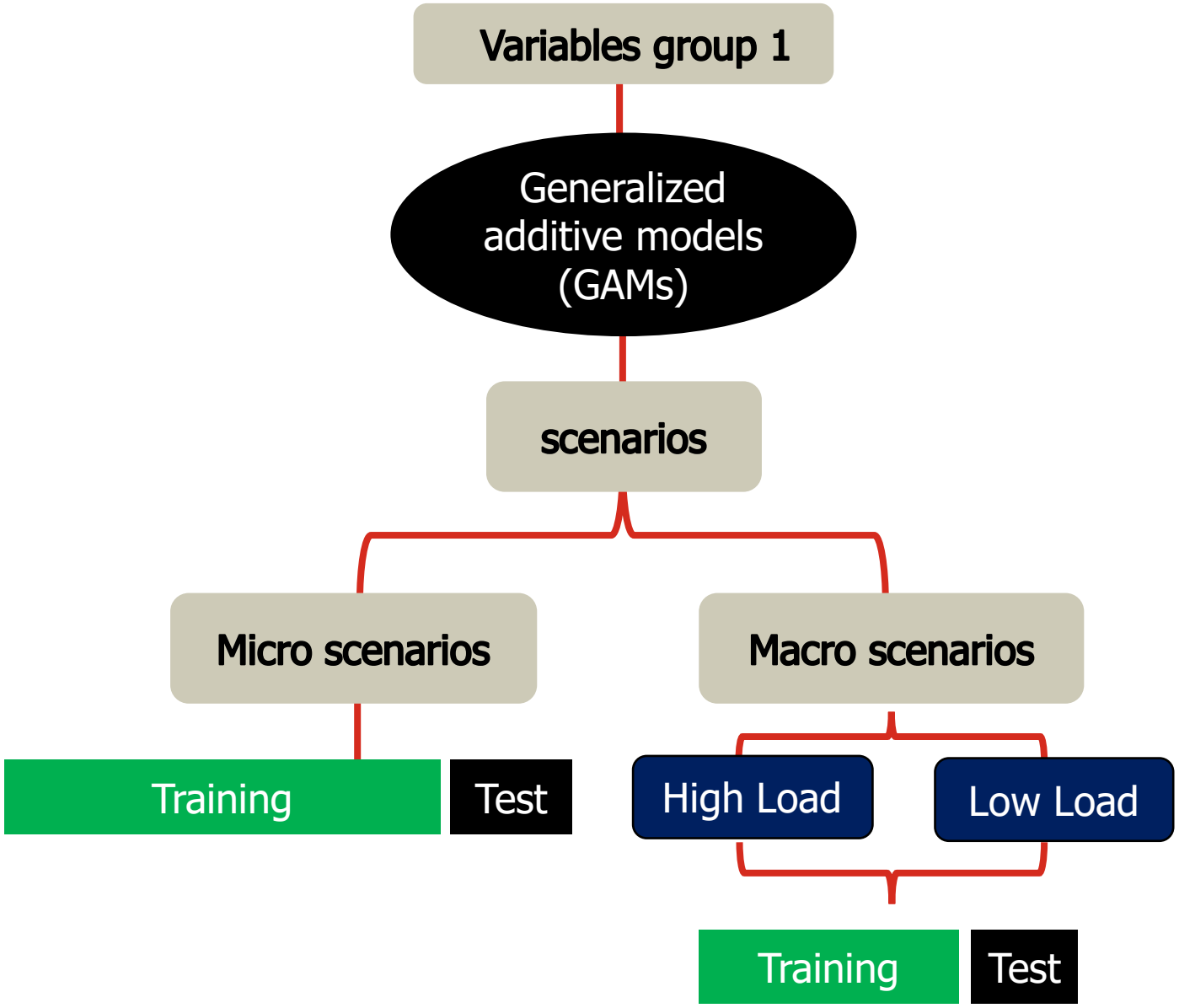
Results obtained from Lasso Models



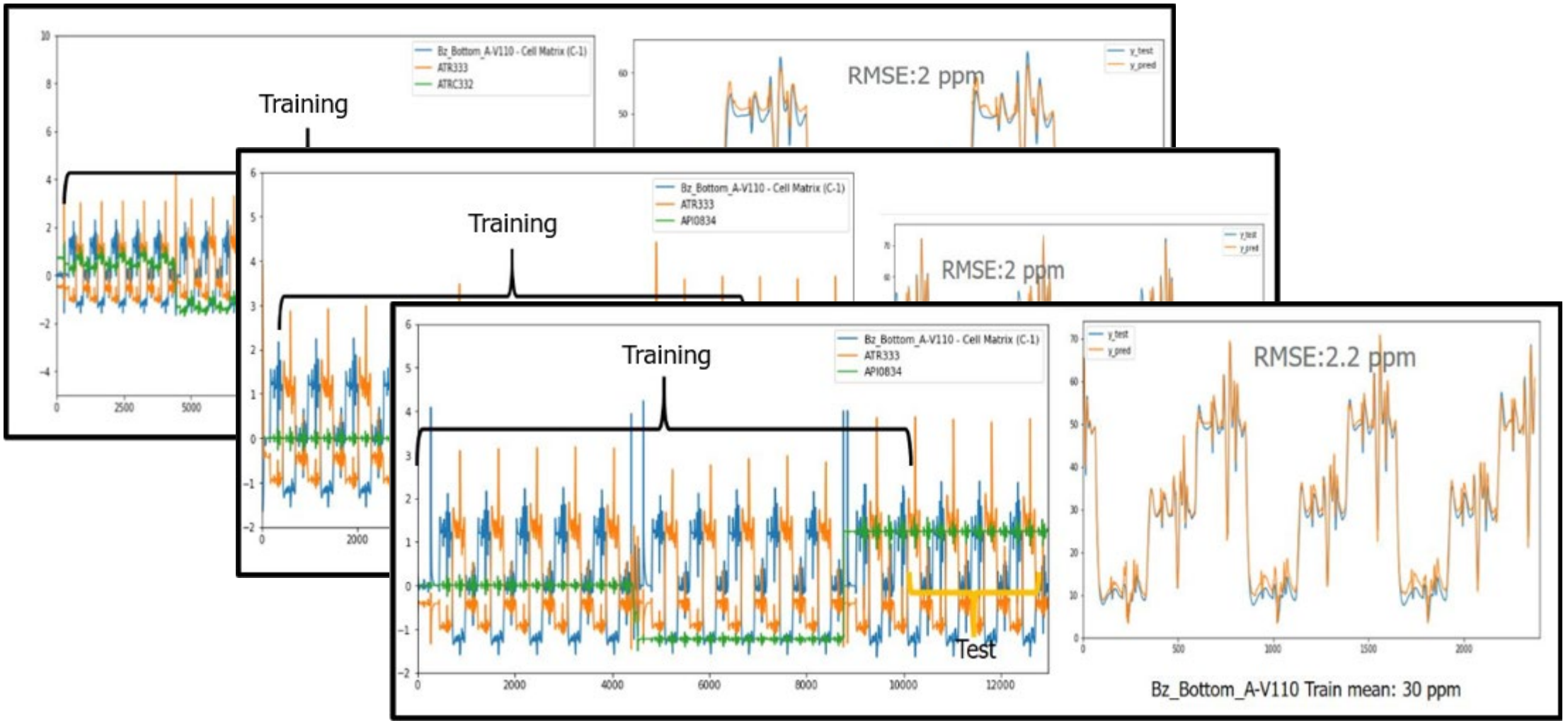
Better results obtained from GAM Models



Model Testing



Model Testing



Conclusion

- 1.5% energy saving on each column involved.
- Quality and recovery improve on the top and bottom streams of the columns.
- Payback around 1 year or less. Minimum investment and good returns.
- Technically viable the data extraction from simulation to create and train machine learning models, even before plants start-up.

Thank you

Q&A