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)

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.

- 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)

Production capacities:

- **250 kt/y** LAB
- **80 kt/y** LABSA
- **400 kt/y** Paraffins
- **100 kt/y** Solvents

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
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.
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.

Problems experienced on previously executed YET projects:
- Phenol 3 (CQP) – Fractionation train modeling: lack of data
- CQPM Paraffins project - C\textsubscript{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.
Inprocess in brief

**independent** from any ICSS or simulation software provider

**founded in Barcelona by domain experts**

**2006**

**worldwide presence**

**48 countries**

**years experience**

**250+**

**executed projects**

**400+**

**simulation engineers**

**50+**

**training courses**

**320+**

**Keen to share its knowledge with clients**

**Mission**: Deliver the value of Process Digital Twins to achieve Operations Excellence
Our Services and Products

Lifecycle Modelling
- Feasibility studies: Selection of alternatives
- Dynamic simulation studies before plant construction
- Validation of control philosophy
- Operational procedures development/enhancement
- Process train: Ensured 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
- Process simulation courses
- Technology courses
- Knowledge Improvement Program – KIP
- Training for plant operators / technicians

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

Applications & Software Development
- Inprocess Infrastructure Suite
- IPSA: P&ID’s database
- ITOP: Inprocess Training & Operations
- ICCM: Inprocess Competence Management System
- IFLOW: to link process simulators with OILGA
- IPSA: Pressure Swing Absorption 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

Operator Station:
  - IIS for Operator Station

Instructor Station and Simulation Server:
  - inprocess infrastructure Suite(IIS) for Instructor Station
  - Aspen HYSYS® Dynamics Run-time

Virtual Reality Station:
  - High-End Nvidia graphic card
  - Virtual Reality Headset
Scope of HYSYS model of the OTS

OTS with 3D VR

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

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.
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.

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 decisions 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 programming skills

**Note:** Aspen Hybrid Models were not available at the time of the project.
Data discovery and analytics

Descriptive Statistic

Descriptive statistics

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 overfitting

Variables group

The groups were built based on the importance of the variables and the knowledge 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
- ATRC332
- ATR134

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

Keep in Outliers

Preprocessing data → Grid Search → Fit → Metrics

Keep Out Outliers

Preprocessing data → Remove Outliers → Grid Search → Fit → Metrics
Machine learning modeling

- Variables groups 1 & 2
  - Generalized additive models (GAMs)
  - Linear regression models (Lasso)
  - Keep out Outliers
  - General models with all data
Results obtained from Lasso Models
Better results obtained from GAM Models

Models generated with all data

GAMs

Group 1

Group 2

Group 3

Group 4
Model Testing

Variables group 1

Generalized additive models (GAMs)

scenarios

Micro scenarios

Training

Test

Macro scenarios

High Load

Low Load

Training

Test
Model Testing

- RMSE: 2 ppm
- RMSE: 2.2 ppm

Training

Test
• 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