Modelling of aluminum reduction processes with machine learning approaches and mathematical models using SAP HANA in-memory computing for smart-devices

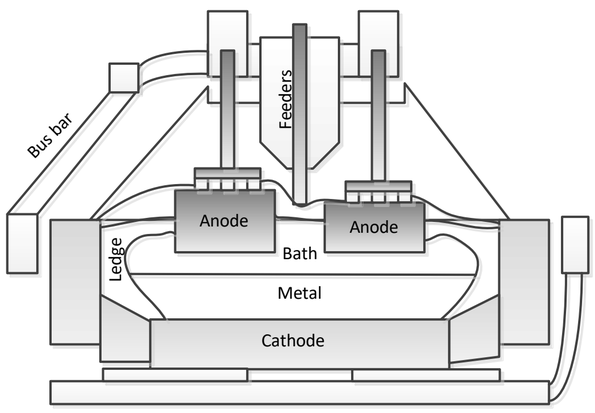
Prof. PhD Roberto C. L. de Oliveira
Post-Graduation Program in Electrical Engineering
Federal University of Pará, Brazil
limao@ufpa.br

Fábio M. Soares
Post-Graduation Program in Electrical Engineering
Federal University of Pará, Brazil
fms@ufpa.br

Project idea

Metallurgical processes are usually very complex and multidisciplinary, resulting in hard challenging control strategies. Process modeling is a very useful approach to understand, investigate and simulate processes' phenomena. Due to the complexity, a pure and physical modeling (white box) can be very time-consuming and not represent the real process as expected, however data-driven modeling (black box) requires less-effort and provides satisfactory results. On the other hand, black box models are purely based on a plant's history, while white box models include every possible scenario. Taking both paradigms together, a stronger model can be built, as with white box models complementary data can be generated for data-driven modeling, which are more simple and can be implemented in smart devices.

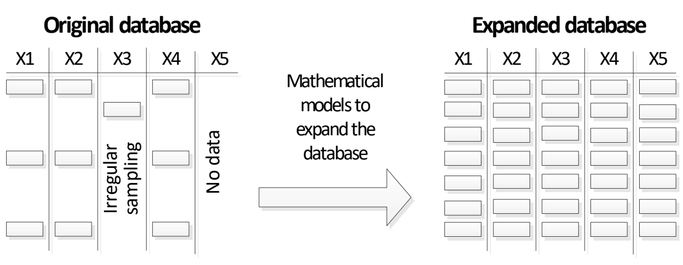
This approach is applied on the Hall-Heroult process for aluminum smelting. A diagram of a typical electrolysis cell is shown below.



A high amperage electric current crosses the cell from the anode to the cathode, reducing alumina into aluminum in a electrolytic bath at temperatures of around 960o C. Alumina is continually added from the feeders into the bath, in order to keep the production rate as high as possible.

Liquid aluminum falls to the bottom of the cell due to gravity and electrical atraction to the cathode, and is periodically tapped. A solid layer called ledge is formed around the bath, thereby providing a protection to the cell walls. A high production rate depends on a stable chemical composition of the bath, temperature and cell resistance. However, such operational events like anode change cause disturbances in the thermal and energy balance, and the control systems must preserve the production as high as possible.

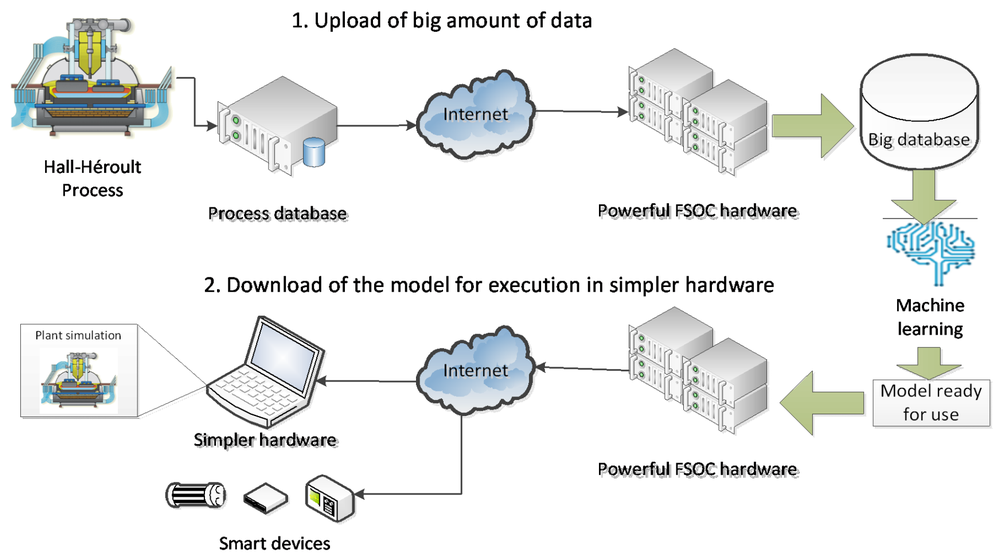
In order to be a viable industry, aluminum smelters must have hundreds of production cells in series (potlines), since a very high electric current must be used to reduce alumina into aluminum. Moreover, the number of variables to monitor may be as high as hundreds, therefore one must pay attention to each of the variables for each of the cells. A typical year of operation of a potline with daily records has about 200 MB, but if this database can be extended by mathematical models, the size can be raised to about 4 GB. The database extension can be made using theoretical models, available in the literature. A simplified schema of database expansion is shown below:



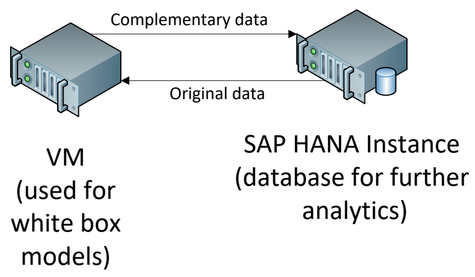
The original database contained records sampled at 24, 32 and 64 hours, by using mathematical models and constant for the type of cell used, the expanded database has a sampling time of 4 hours. About 60 variables are included in the model, from which 15 are original and 45 are generated. Backpropagation Neural Networks are chosen as the machine learning structure, whereby the hidden layer size is varied from 1 to 100 hidden neurons, to find the simpler yet with good generalization. In addition, we performed dimensionality reduction with Principal Components Analysis (PCA) before training the neural networks.

Used FSOC resources

For this project the general scenario shown in figure below was designed.

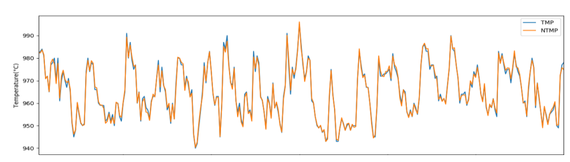


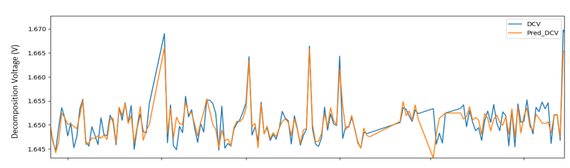
This project used eight virtual machines (VMs), four extra large (XL) for cell data manipulation (960 cells) and four large (L) for potline data manipulation (4 potlines), a scenario similar to a smelter. In these VMs the theoretical models are run to produce the complementary data. The original database is uploaded into a SAP HANA tenant for faster querying and processing. The VMs retrieve the original data, run the models and after producing the complementary data, they are stored in the SAP HANA database.

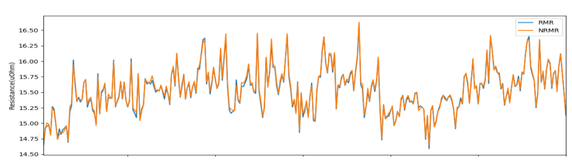


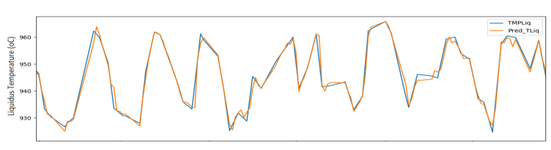
Preliminary findings and next steps

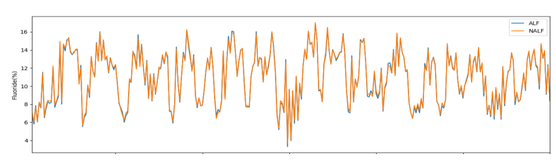
Below are the charts for the variables modeled so far, the actual data are in blue and the neural models are in orange.

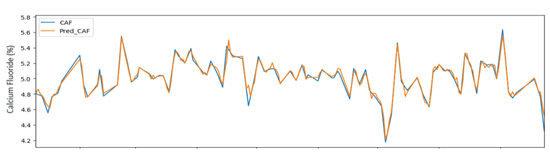

Bath Temperature (TMP)

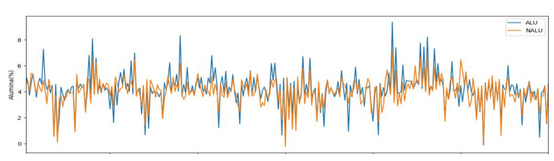

Decomposition Voltage (DCV)

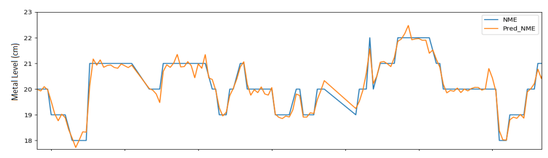

Cell Resistance (RMR)


Liquidus Temperature (Tliq)

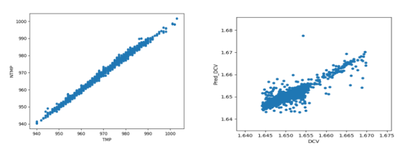

Aluminum Fluoride (ALF)


Calcium Fluoride (CAF)

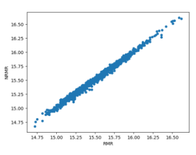

Alumina (ALU)

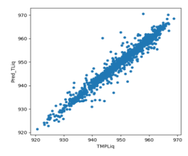

Metal Level (NME)

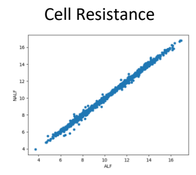
Below are the scatter charts for the predictions of all variables

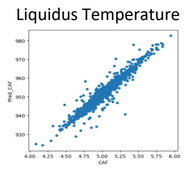

Bath Temperature

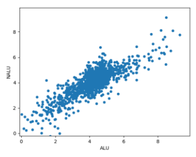
Decomposition Voltage

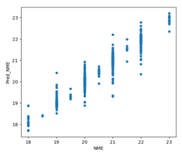

Cell Resistance


Liquidus Temperature


Aluminum Fluoride


Calcium Fluoride


Alumina


Metal Level

With the exception of Alumina and the Decomposition Voltage, which are not present in the original database, but are generated through models, the predictions are farily good. It is worth mentioning that these results are obtained with a few number of neurons, what would facilitate the task of implementing on cheaper hardware. Below are the summary of the neural models and machine learning task.

Machine learning summary

Total dataset	250000 records
Testing dataset	15%
Production cells used	24 cells from the same potline
Period of operation	6,5 years

Neural Models Summary

Variable	Hidden Neurons	MSE Train Error	MSE Test Error	Prediction Correlation
TMP	66	0.00078	0.001083	0.9962
RMR	61	0.000668	0.00128	0.9918
ALF	42	0.000961	0.001815	0.9945
ALU	81	0.002315	0.004563	0.9277
DCV	36	0.001539	0.002624	0.8975
CAF	31	0.00178	0.002932	0.9635
NME	27	0.001168	0.002216	0.9744
Tliq	30	0.001842	0.003025	0.98

To improve the results, we will refer to the new studies on aluminum reduction process modeling and use theoretical spatial data information from partial differential equations, that would require a massive processing unit, such as GPU. In addition, other techniques for validation should be tested.