Computational Intelligence for Sustainable Glass Manufacturing: A Data-Driven Approach for Energy Efficient Conditioning

David PeÃ±a-Mangas*[1], Carlos Cernuda[1], and Daniel Reguera-Bakhache[1]

[1]Faculty of Engineering, Electronics and Computing, Mondragon Unibertsitatea
Arrasate-Mondragon, Spain

Abstract

In the glass container manufacturing process, conditioning is a key stage that contributes to energy consumption. The main objective of conditioning is to cool the glass exiting the furnace to a suitable temperature for container forming. Currently, this stage is managed based on the experience of operators, which is functional but not optimized for energy efficiency. While several approaches to minimize consumption based on process control using physical modeling have been proposed in the literature, they do not completely account for all the involved variables. Moreover, none of these studies leverage the power of data to predict energy consumption patterns. In this paper, we introduce a data-driven method to minimize energy consumption during the glass conditioning stage. We applied this methodology to a specific production line and tested it under various scenarios, achieving potential absolute savings up to 10% in energy consumption. Operational validations in two additional real forehearths showed energy reductions of 26.3–89.3 kWh per operating hour, corresponding to relative savings of 8.2–22.1%, including a same-production A/B test. The implementation of this method has the potential to significantly contribute to the decarbonization goals of the glass manufacturing industry.

1 Introduction and related work

The glass manufacturing industry faces increasing pressure to enhance both sustainability and operational efficiency. Traditional control methods â€”often based on operator experienceâ€” are reaching their limits in addressing the environmental challenges inherent in high-energy processes such as glass container manufacturing. In an era where reducing the carbon footprint and optimizing resource use have become imperative, there is a critical need for innovative approaches that integrate data-driven techniques with process optimization.

Inside a glass container factory, almost every aspect of the process is crucial to obtain a zero-defect container, which complies with the quality requirements of the customer. This process, as shown in Figure 1, includes: (i) melting raw materials (mainly silica sand, soda ash, and lime) at over 1500 ÂºC in large furnaces; (ii) glass conditioning in the forehearth; (iii) gob-forming in the feeder; (iv) container forming in the IS (Individual Section) machine and its respective molds; and (v) container transportation and inspection.

*david.penam@alumni.mondragon.edu

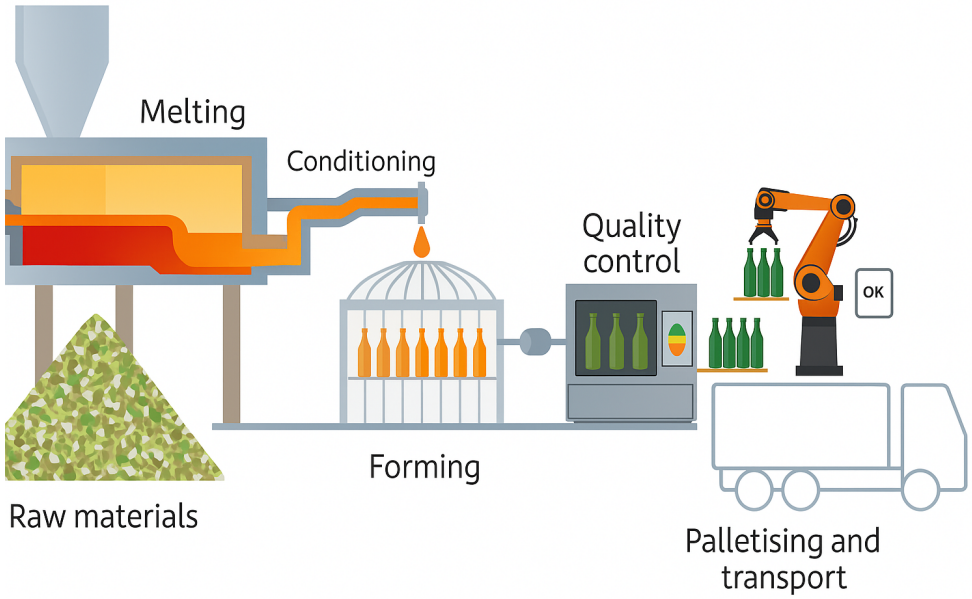

Figure 1: Glass container manufacturing line typical layout.

The main task of the conditioning process is to cool the glass to achieve the desired temperature for manipulation and gob forming. This temperature, depending on some factors such as the type of forming process (Press & Blow or Blow & Blow), the glass color or the type of container to be produced, varies between 1150ÂºC and 1250ÂºC. A typical conditioning system consists of a distributor (or Working End) at the outlet of the furnace, and forehearths (longitudinal channels), each of which ends in a feeder with a corresponding spout bowl where the glass gob is formed. This layout is shown in Figure 2.

Glass conditioning requires a series of heating and cooling systems. Despite cooling being the main task of the conditioning process, a heating system is necessary to accurately control the temperatures. This is to prevent unwanted heat losses[1], which may provoke thermal inhomogeneity within the glass bath. The forehearth is divided into several zones along the length, each of which with a single or various temperature setpoints. Automatic temperature controllers are responsible for maintaining the process values as near as possible to these setpoints by means of heating and cooling mechanisms.

The central flow of the forehearth maintains a higher temperature, whilst the side flow undergoes wall heat losses that make the temperature decrease. To compensate this natural scenario, the forehearth is usually also divided into cooling and heating channels across the transverse axis[1], as shown in Figure 3.

When the glass reaches the spout bowl, it is formed into gobs. At this point, three triple thermocouples control the thermal homogeneity of the glass to achieve a sequence of uniform temperatures. Several formulas have been proposed to calculate a homogeneity key process indicator[2].

One of the key aspects that most determines the efficiency of the forehearths is the conditioning operation. Related literature in this area is scant, and knowledge has to be extracted either from experienced operators or from information and articles on the web. However, some physical modeling works have been published, which help to understand the

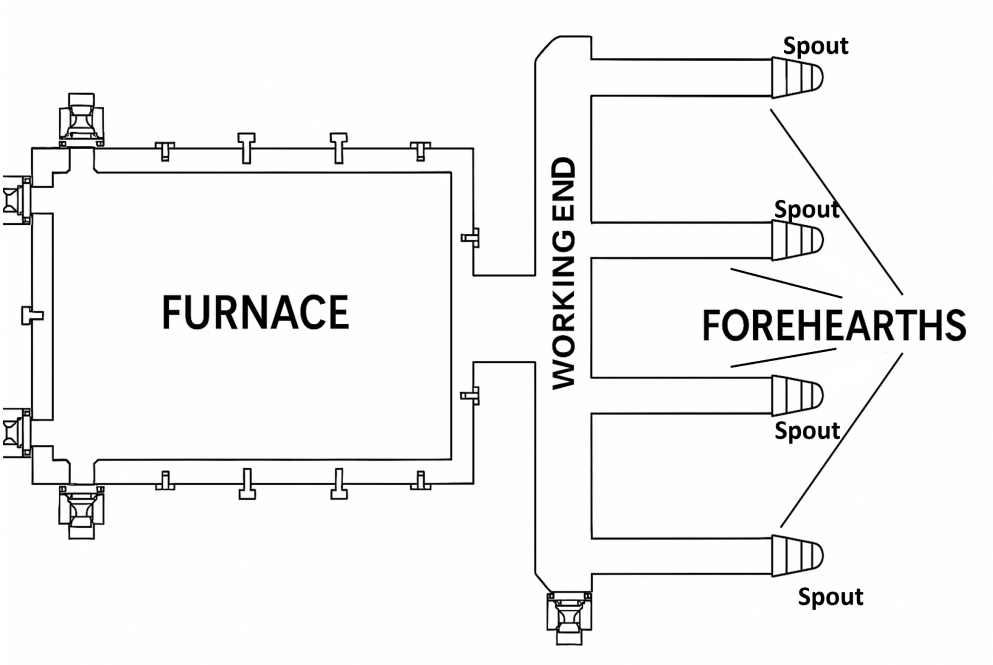

Figure 2: Typical furnace, working end, and forehearths layout in a top view.

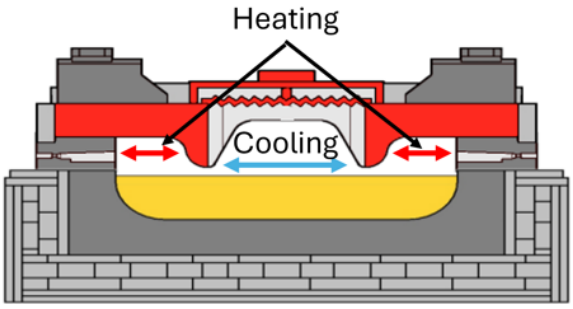

Figure 3: Typical Forehearth layout. Cross-sectional view.

phenomena occurring inside the forehearths. The most complete simulation was carried out by Hyre and Paul in[3], where they constructed a three-dimensional model for flow and heat transfer in working end/forehearth systems. The simulation included both the combustion space above the glass and the cooling airflow along the centreline of the forehearth.

These kinds of simulations assist the development of process control solutions. In addition to traditional Proportionalâ€“Integralâ€“Derivative (PID) controllers, some other advanced controllers such as the cascade secondary loop[4] or the predictive adaptive controllers[5] have been developed. Feedforward-feedback control loops based on Partial Differential Equations (PDE) have also been simulated, such as those proposed by Kharitonov et al.[6;7], which proved successful in managing temperature changes. More recently, Drapala et al. developed Model Predictive Control (MPC) algorithms for the stabilization of forehearth temperatures[8;9], and Linear Quadratic Regulator (LQR) based control[10], which provided better results than MPC.

Despite the advanced state of process control and modeling reviewed above, there exists a general lack of knowledge regarding energy savings in glass conditioning. Henkel et al.[11] developed a steady-state energy optimization based on the one-dimensional model described before[7]. It consisted of the modification of intermediate temperature setpoints to achieve the overall minimum model output, where the output is gas consumption. Despite the promising nature of this approach, it does not leverage the power of historical data to develop the consumption model. Beyond the glass sector, some related work has been carried out in similar manufacturing processes such as steel production. Vesel and Isaacs[12] created a control system optimizer using an open-architecture graphical programming software toolkit with deep learning capabilities for a steel reheat furnace. Nevertheless, this reheating process, despite being somewhat similar to the glass conditioning, does not involve any cooling system and is a solid material at that point of the process, which makes the heat transfer totally different from the case of glass, which can be considered a fluid at the conditioning stage.

In recent years, Data-Driven approaches have gained relevance in the field of energy optimization across various industrial processes. For example, in the steel industry, Machine Learning techniques such as Decision Trees and Multilayer Perceptrons have been applied to predict energy consumption in electric arc furnaces, enabling more efficient operational planning[13]. Similarly, in the food industry, neural networks have been successfully used to forecast energy demand in cogeneration plants, aligning production schedules with energy availability and significantly reducing CO_2 emissions[14]. In the automotive sector, the combination of high-fidelity simulations and AI-based reduced-order models, such as variational autoencoders, has been employed to optimize the thermal curing process of car body paint, achieving substantial energy savings while preserving process quality[15]. These examples highlight the growing potential of data-driven techniques for energy optimization in industrial environments. However, despite the energy-intensive nature of glass conditioning, this process remains largely unexplored from a data-driven optimization perspective. Therefore, this is one of the glass manufacturing process parts in which a larger knowledge gap exists. To the best of our knowledge, there is currently no data-driven system available that can minimize the energy consumption of glass conditioning.

This paper proposes a data-driven methodology to minimize the energy consumption of the glass conditioning process in a glass container manufacturing line. Taking into consideration the production conditions of the process, the proposed method provides suitable intermediate temperature setpoints in the forehearth, which minimize the amount of energy required. In addition to the model-based evaluation, the methodology was validated in two additional real forehearths: one by comparing optimized operation against previous production series of the same production jobs, and another through a same-production A/B test after changing the recommended setpoints during operation.

2 Methodology

The energy consumption of the glass conditioning stage is almost exclusively due to the natural gas consumption for the glass heating. The gas consumption is determined by the air/gas mix pressures of the burners of all the zones in the working end and forehearth. We try to minimize the sum of these mix pressure values. To achieve it, the first task is to be able to predict the mix pressure of each zone based on the distributor and forehearth conditions. The conditions that affect to the energy consumption are:

- **Glass Residence time.** How long the glass takes to pass through the forehearth. Usually between 40 and 120

minutes. More residence time in the distributor and forehearths implies higher temperature losses and therefore more energy input to maintain the desired temperatures.

- **Glass heat transfer.** Heat transfer is directly affected by the glass composition or color and by the glass bath geometry. This geometry is defined by factors such as the glass level or glass bath width, directly affecting the heat transfer of the glass both internally (conduction and convection) and externally (radiation) to the atmosphere and forehearth walls.
- **Required cooling.** More required cooling makes it more efficient to achieve the desired temperature by applying less heating. On the other hand, less required cooling means that more heat must be applied to maintain the desired temperatures.

If these factors are translated to accurately measured variables that determine the previous ones, the following ones are obtained:

- **Pull.** A higher pull makes the glass to flow faster, reducing the residence time.
- **Forehearth geometry.** The geometry affects the glass heat transfer, as mentioned before, and also the glass residence time, since a bigger glass bath total volume implies more time for the glass melt to arrive the spout.
- **Furnace exit temperature.** Directly proportional to required cooling. Hereafter named TR, as usually done within the industry.
- **Forming temperature.** Inversely proportional to required cooling.
- **Conditioning operation.** Intermediate temperature setpoints and heating/cooling controllers' parameters.

The last of the 5 factors is the one that, given certain conditions in the furnace and production (the ones that define the other 4 factors, and which are fixed by the melting and forming teams), is to some extent editable by the workforce responsible for the conditioning. That factor is mainly configurable through the intermediate temperature setpoints, since the other parameters are set constant for stability reasons. Therefore, the main goal of this work is to develop an optimizer that, given the first 4 factors, provides suitable intermediate temperature setpoints, which minimize the required amount of energy for the conditioning process. Additionally, the resulting setpoints must comply with some restrictions such as a favorable thermal homogeneity, calculated as in the literature, that was described in Section 1.

To this end, a 2-stage methodology is proposed:

1. Create a mathematical model that is able to predict the energy consumption based on the 5 variables described above.
2. Run an optimization algorithm that, given the input conditions, is able to get the most advantageous temperature setpoints in terms of energy efficiency, while maintining process control and high thermal homogeneity.

It should be clarified that the purpose of this work is not to develop a new control algorithm for forehearth temperature stabilization, but to demonstrate how data-driven models can be effectively applied in industrial glass conditioning environments. The existing industrial control loops remain responsible for maintaining the glass temperature around

the imposed setpoints. The proposed method operates at a supervisory level: it recommends alternative intermediate temperature setpoints that are expected to reduce energy consumption, while predictive supervisory models penalize combinations that would lead to unrealistic or undesirable operating conditions.

2.1 Consumption prediction models

As mentioned before, the proposed approach is to predict the air-gas mix pressure of each zone, as a function of: line specific pull, furnace exit temperature, forming temperature and conditioning intermediate temperature setpoints. The pressure values are always between 5 and 65 mbar, to avoid backfires and/or surface overheating. They were normalized to 0-1 variables by applying the function: $MIX' = (MIX - 5)/60$. Additionally, the amount of cooling also plays key role in these predictions, since more cooling requieres more heating to compensate, due to heat transfer by conduction inside the glass melt. The cooling values are measured through the PID controller output, which is a percentage value between 0 and 1.

For ease of notation, for each zone i, the variables are named as in Table 1.

Table 1: Variables notation for each zone i.

Name	Description	Units	Range
$Side\_SP_i$	Temperature setpoint at side channel.	ÂºC	1150-1400
$Side\_PV_i$	Temperature process value at side channel.	ÂºC	1150-1400
$Center\_SP_i$	Temperature setpoint at center channel.	ÂºC	1150-1400
$Center\_PV_i$	Temperature process value at center channel.	ÂºC	1150-1400
MIX_i	Air-gas mix pressure.	mbar	5-65
$Cool_i$	Cooling controller output.	%	0-1

To avoid overfitting, an individualized approach was followed, i.e. we trained a model for each zone, where the output is the air-gas mix pressure of that zone and the input variables are only the indispensable ones. Given a zone i, where the model output is MIX_i, the set of variables that affect this output, and that is used as the model input is:
$\mathcal{S}_i$:= {$Side\_SP_i$, $Side\_PV_{i-1}$, MIX_{i-1}, $Cool_i$, Pull}
and the corresponding model is named M_iMIX. With this baseline, some annotations must be done:

- A sequential approach must be followed, since the output of zone $i-1$ (MIX_{i-1}) is used as input for the zone i.
- $Cool_i$ is also unknown from initial conditions, so a separate model is trained to predict it, and added to the sequence of models.
- In case $i = 0$, the zone $i-1$ would be the furnace exit or a previous working end zone that is not part of the line consumption. MIX_{i-1} would be used only if controlled automatically, not if it is set constant.
- Due to some of these temperatures having low variability in the dataset and to further avoid overfitting and bad generalizations, $Side\_PV_{i-1} - Side\_SP_i$ is taken as input, reducing the number of input variables by one. This new variables are named $Side\_Diff\_SP_i$ and are included in $\mathcal{S}_i$.

Thus, as mentioned in the second point above, a separate model for the cooling must be trained for each zone where cooling exists. Hence, the input variables for the models are detailed in Table 2, where $Center\_Diff\_SP_i =$

$Center\_PV_{i-1} - Center\_SP_i$. Note that for $i = 0$ the variable $Cool_{i-1}$ is only used if controlled automatically, as for the MIX model.

Table 2: Variables and notation for primary models.

Channel	Model name	Input variables	Input notation	Output
Center (C)	M_iCool	$\{Center\_Diff\_SP_i, Cool_{i-1}$, Pull$\}$	$\mathcal{C}_i$	$Cool_i$
Side (S)	M_iMIX	$\{Side\_Diff\_SP_i, MIX_{i-1}, Cool_i$, Pull$\}$	$\mathcal{S}_i$	MIX_i

With the same objective of avoiding overfitting, the models to train must have as few parameters as necessary to adjust. Being the output variables within the range [0-1] with asymptotic nature in its extremes, the best results were obtained with a sigmoid neuron. The sigmoid neuron model fits the Equation 1:

$$y = \frac{1}{1 + \exp\left(-\left(\beta_0 + \sum_{j=1}^{n} \beta_j x_j\right)\right)} \tag{1}$$

Therefore, the term data-driven model should not be interpreted here as a deep neural-network architecture. The primary prediction models can be understood as sigmoidal regression models: the measured process variables are combined linearly through the coefficients β_j and then mapped to one normalized output by means of a sigmoid function, without hidden layers. This deliberately simple structure was selected to keep the models interpretable for process engineers, to reduce overfitting risk in industrial datasets with limited variability, and to make the optimizer behavior easier to audit. The models were implemented and trained in Python using the Keras environment; due to the small number of adjustable coefficients in each zone-specific model, the numerical training stage is lightweight compared with the selection, cleaning, and validation of representative plant data.

2.2 Optimization Methodology

After achieving a correct prediction of the gas energy consumption, an optimization algorithm is executed. For that, an objective function has to be defined previously, that is composed of the sum of the obtained outputs by the models described, and some penalizations to avoid unwanted optimizations. Mainly, we do not want the algorithm to provide:

- **Unrealistic temperature setpoints.** The algorithm could provide setpoints that are theoretically efficient in terms of energy consumption, but that are not achievable by the system, what creates high instability and risk in the forehearth. For example, it could provide a 100Â ºC difference between both setpoints of the same zone, the one of the center and the one of the side, what is completely absurd. This penalty is named *Physics Loss*, as substantial deviations indicate that the suggested setpoints are physically unattainable.
- **Low homogenity.** One of the main tasks of the glass conditioning process is to provide the forming machine a thermally uniform glass melt. Therefore, we try to avoid energy-efficient configurations that hinder the forming process.

To compute these penalizations, supervisory surrogate models are trained.

2.2.1 Physics loss (PL)

In this case, surrogate models for predicting PV_i are trained for both side and center channels, based on the predicted output of the primary models M_iMIX and M_iCool. This is done to compare the predictions to the setpoints that are set. The penalization is computed as in Equation 2 for any zone i:

$$\text{Physics loss} = g(|SP_i - \overline{PV_i}|) \tag{2}$$

Being $\overline{PV_i}$ the prediction of PV_i by the corresponding surrogate model and g any non-decreasing monotonic function to adjust the values for penalization. The input for those supervisory models are detailed in Table 3. These supervisory models are linear. Moreover, with the same logic as in the primary models to avoid overfitting, instead of using PV_{i-1} as input and PV_i as output, $Diff\_PV_i = PV_{i-1} - PV_i$ is used as output, reducing the number of input variables by one –first element in $\mathcal{C}2_i$ or $\mathcal{S}2_i$ can be removed–.

Table 3: Variables and notation for supervisory surrogate models predicting PV_i. Variables $Cool_i$ and MIX_i are translated to $MiCool(\mathcal{C}_i)$ and $M_iMIX(\mathcal{S}_i)$, respectively, in the last column. The rest of the variables are grouped into $\mathcal{C}2_i$ and $\mathcal{S}2_i$, respectively.

Channel	Model name	Input variables	Input notation	Output
Center (C)	$M2_iCenter\_PV$	$\{Center\_PV_{i-1}, Cool_{i-1}, MIX_{i-1}, \text{Pull}, Cool_i\}$	$\{\mathcal{C}2_i, MiCool(\mathcal{C}_i)\}$	$Center\_PV_i$
Side (S)	$M2_iSide\_PV$	$\{Side\_PV_{i-1}, Cool_i, \text{Pull}, MIX_i\}$	$\{\mathcal{S}2_i, M_iMIX(\mathcal{S}_i)\}$	$Side\_PV_i$

These surrogate models are linear regression models. This choice is consistent with their supervisory role: they are not intended to reproduce the full thermal dynamics of the forehearth, but to provide a simple and interpretable estimate of whether a candidate setpoint is physically coherent with the process behavior observed in the historical data.

These models act as supervisors of the input setpoint temperatures. Taking for example a side channel, the proposed approach is depicted in Figure 4. Given the primary model prediction MIX_i or $Cool_i$, and the variables in $\mathcal{S}2_i$ or $\mathcal{C}2_i$, respectively, we get the complete input of $M2_iSide\_PV$ or $M2_iCenter\_PV$. Thus, it can be predicted $Side\_PV_i$ or $Center\_PV_i$, respectively. The distance between this output and $Side\_SP_i$ or $Center\_SP_i$ can be considered as a measurement of unrealistic setpoints. If this difference is too high, it means that the given setpoints are not attainable. Therefore, the physics loss penalization of Equation 2 is rewritten as in Equation 3 for side and center channels.

$$\begin{aligned} \text{PL\_}S_i &= g\Big[\big|Side\_SP_i - M2_iSide\_PV(\mathcal{S}2_i, M_iMIX(\mathcal{S}_i))\big|\Big] \\ \text{PL\_}C_i &= g\Big[\big|Center\_SP_i - M2_iCenter\_PV(\mathcal{C}2_i, M_iCool(\mathcal{C}_i))\big|\Big] \end{aligned} \tag{3}$$

2.2.2 Homogenity loss (HL)

For the case of homogenity penalization, an homogenity prediction model is trained based on the forehearth temperatures. However, since various forehearth temperatures are configurable (usually over 5), any model is able to get overfitting to predict the homogenity. Therefore, we tried to be accurate in predicting all the 9 temperatures measured in the spout through linear models, with the minimum inputs for each, and then compute the homogenity value based

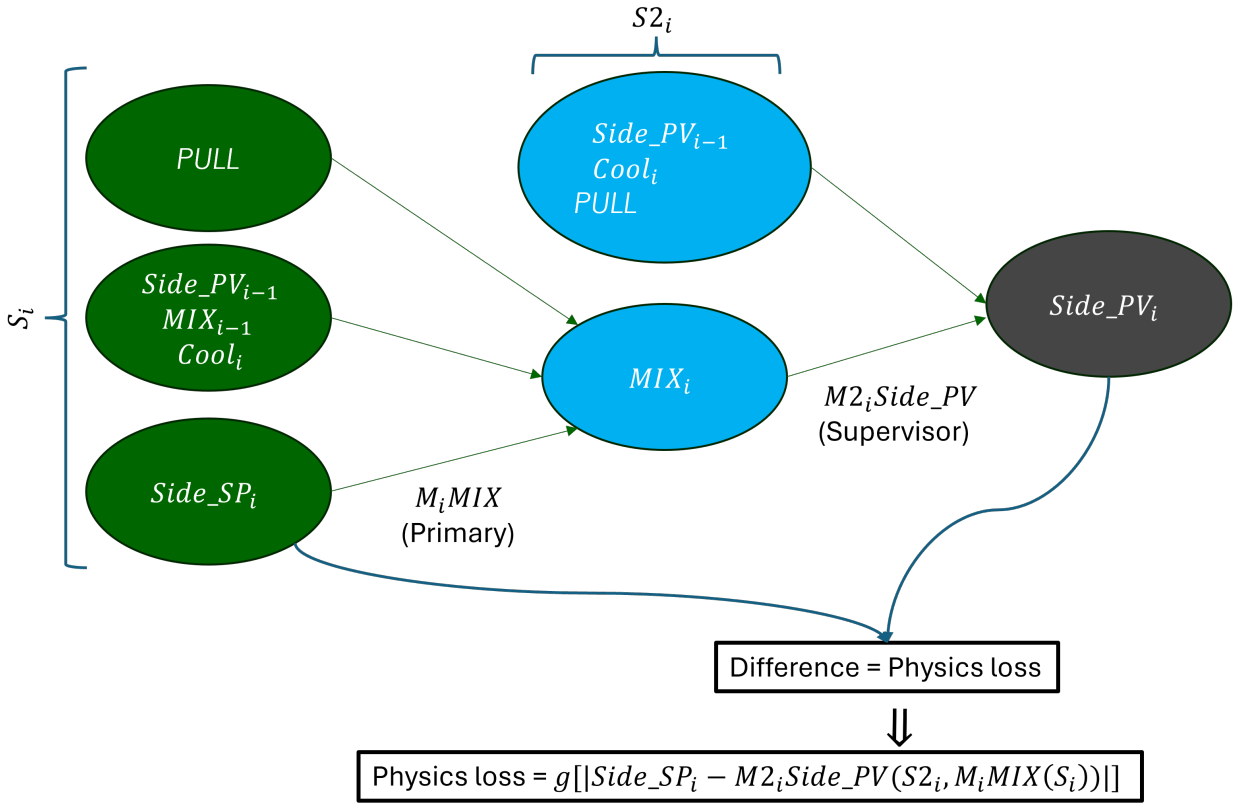

Figure 4: Proposed approach. The primary model prediction is supervised by means of a secondary surrogate model that predicts the otput temperature, which is compared to its setpoint.

on those predictions. To the final prediction, similarly to the first penalty, a non-increasing monotonic adjusting function h is applied. The inputs used for each temperature readings are shown in Table 4, where n is the number of zones of the forehearth. These inputs are based on process expertise, since bottom temperatures are affected by the initial temperatures of the forehearth, and the higher up, the more important are the temperatures closer to the spout.

The resulting model is named $M\_Homogenity$, and the penalty is computed as:

$$\text{HL} = h(M\_Homogeneity(input)) \tag{4}$$

2.2.3 Resulting objective function

Once described how the penalties are applied to the optimization, the final objective function is:

$$f(SP) = \text{EC} + \text{PL} + \text{HL} \tag{5}$$

Table 4: Input temperatures used to predict the 9 temperature values in the spout. TC refers to the thermocouple locations in the 3x3 matrix of thermocouples

Zone	**n**	**n-1**		**n-2**		**n-3**	
Channel / **TC**	**Side**	**Side**	**Center**	**Side**	**Center**	**Side**	**Center**
Center top	✓		✓				
Sides top	✓						
Center middle	✓		✓		✓		✓
Sides middle		✓		✓	✓		✓
Center bottom			✓		✓		✓
Sides bottom		✓		✓	✓		✓

where EC refers to energy consumption. Each term of the objective function is rewritten below.

$$\begin{aligned} \text{EC} &= \sum_{i=1}^{n} \frac{M_i MIX(\mathcal{S}_i)}{n} \\ \text{PL} &= \sum_{i=1}^{n} (\text{PL\_}C_i + \text{PL\_}S_i). \text{ Terms defined in Equation 3.} \\ \text{HL} &\text{ defined in Equation 4} \end{aligned} \tag{6}$$

where n refers to the number of zones. That is, the average air-gas mix pressure of the zones (EC. Value within the range [0,1]) plus the sum of all penalizations.

2.2.4 Optimization procedure

Once defined the objective function, an optimization algorithm capable of searching a global minimum is taken from the literature. Powellâ€™s method[16] was selected for its proven efficiency and robust performance in minimizing differentiable and non-differentiable functions without requiring gradient information. This derivative-free characteristic is particularly advantageous when integrating multiple linear predictive models into a single objective function, as computing or approximating gradients can become prohibitively complex. In addition, Powellâ€™s algorithm benefits from efficient one-dimensional line searches along carefully chosen search directions, enabling it to converge to the minimum with fewer function evaluations compared to more general-purpose methods. Consequently, Powellâ€™s method offers both speed and reliability, making it well suited for this application where rapid and stable convergence is crucial.

3 Results and Discussion

The results are analyzed by the accuracy of the prediction models –mix pressure, temperatures and homogenity predictions– as well as the optimization predicted savings in energy consumption. This all was tested in a single production line that was used as use case and that is described below.

3.1 Experimental setup. Use case line.

The research focused on a single production line, where the forehearth geometry variable is obviously constant. The studied line is as follows in terms of conditioning:

1. **One working end zone**, with stepping direct control. This kind of control means that glass temperature is always controlled by gas heating and, if necessary –if the heating goes out of the limits–, cooling output is changed in steps by a PLC until the heating variable is again between the defined bounds.
2. **First forehearth zone**. Same heating controller at both sides and independent direct cooling at the center channel.
3. **Second forehearth zone**. Same heating controller at both sides and independent direct cooling at the center channel.
4. **Third forehearth zone**. Only heating zone. Same heating controller at both sides.
5. **Spout**. Manually controlled heating over the Spout, maintained constant. Is not necessary to be taken into account.

The layout of the forehearth is shown in Figure 5. The forehearth zones 1-3 are named in this way, and to maintain coherence, the Working End zone, which is previous to them, is named zone 0. Therefore, according to this description, the available variables for the prediction models are the ones marked in Table 5.

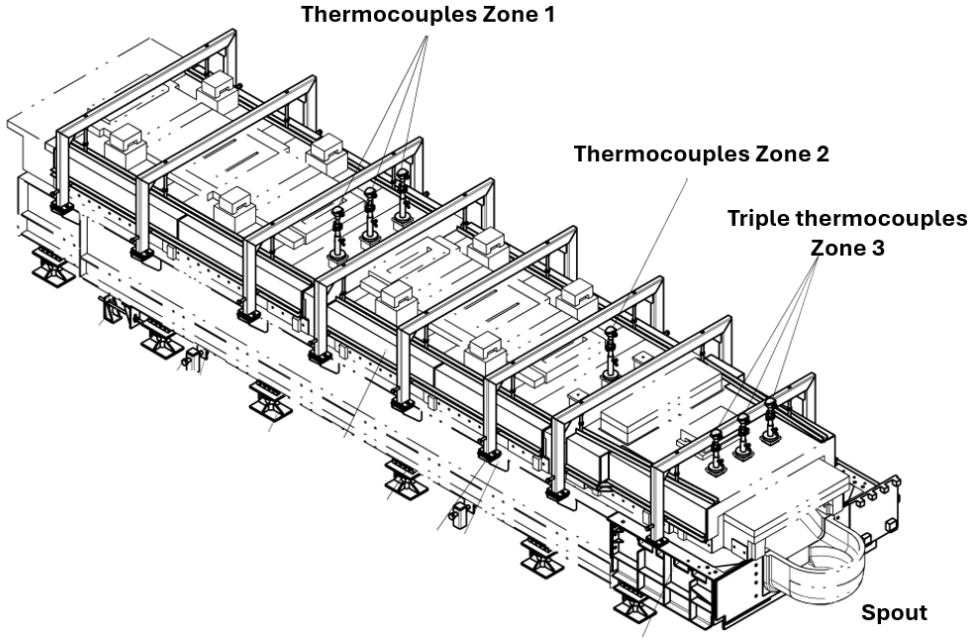

Figure 5: Studied line's Forehearth layout. Burners and cooling air in zones 1 and 2, before the corresponding thermocouples; and only burners in zone 3 (before the Spout) and over the Spout.

Particular observations for the use case:

Table 5: Available variables in the use case line, that have been previously defined in Table 1.

Variable / Zone	$Side\_SP_i$	$Side\_PV_i$	MIX_i	$Center\_SP_i$	$Center\_PV_i$	$Cool_i$	TR	Pull	Color
General							✓	✓	✓
0	✗	✗	✓	✓	✓	✓			
1	✓	✓	✓	✓	✓	✓			
2	✓	✓	✓	✓	✓	✓			
3	✓	✓	✓	✗	✗	✗			

- In case $i = 0$, the zone $i - 1$ is another Working End zone, which is next to the furnace exit. The PV of this temperature is taken as PV_{i-1}, but MIX_{i-1} and $Cool_{i-1}$ are not used since they are constant in this case.
- It must be taken into account when computing the Physics Loss that side and center channel temperature setpoints are not set in all zones. Zone 3 is only heating zone (only side setpoint) and zone 0 has only central temperature, as shown in Table 5.

3.2 Prediction models accuracy

Both primary models (M_iMIX and M_iCool) and surrogate models ($M2_iSide\_PV$, $M2_iCenter\_PV$, and $M\_Homogenity$) are evaluated separately in this section. For that, Mean Absolute Error (MAE) is used as the accuracy metric. It should be noted that in the case of primary models, the output variables are within the range [0,1], whereas in the secondary models, the output variables have their own nature (range [1200,1400] for temperatures and range [$-\infty$,100] for homogenity). All the models were trained for a 3 year period by a train-test split method without shuffling, always assuring that a representative range of available data is used in both training and test sets of each variable, since the variability was limited in some cases. Robust results were obtained with this approach. The data frequency was hourly, having available a total of 21782 dataset rows, of which the 60% was used for training, and 40% was used for test purpose. For the primary models, the results are shown in Table 6.

Table 6: Primary models accuracy.

Output variable	Train/Test	MAE
$Cool_0$	Train	$2.87 \cdot 10^{-2}$
$Cool_0$	Test	$5.17 \cdot 10^{-2}$
MIX_0	Train	$4.17 \cdot 10^{-2}$
MIX_0	Test	$4.07 \cdot 10^{-2}$
$Cool_1$	Train	$1.44 \cdot 10^{-2}$
$Cool_1$	Test	$4.53 \cdot 10^{-2}$
MIX_1	Train	$1.79 \cdot 10^{-2}$
MIX_1	Test	$5.29 \cdot 10^{-2}$
$Cool_2$	Train	$1.65 \cdot 10^{-2}$
$Cool_2$	Test	$1.61 \cdot 10^{-2}$
MIX_2	Train	$2.88 \cdot 10^{-2}$
MIX_2	Test	$4.07 \cdot 10^{-2}$
MIX_3	Train	$3.60 \cdot 10^{-2}$
MIX_3	Test	$12.7 \cdot 10^{-2}$

Despite the accuracy metric being important, it should be borne in mind that the main goal is not to get a perfect

accuracy, but to get the optimal values in each scenario to minimize the consumption. Therefore, the patterns captured by the models are what really matter. The best method to evaluate the models in this way is to look at the predicted vs real values evolution over time. For example, in Figure 6 the model M_1MIX result over time is plotted, whereas in Figure 7 an XY scatter plot of the model result is plotted.

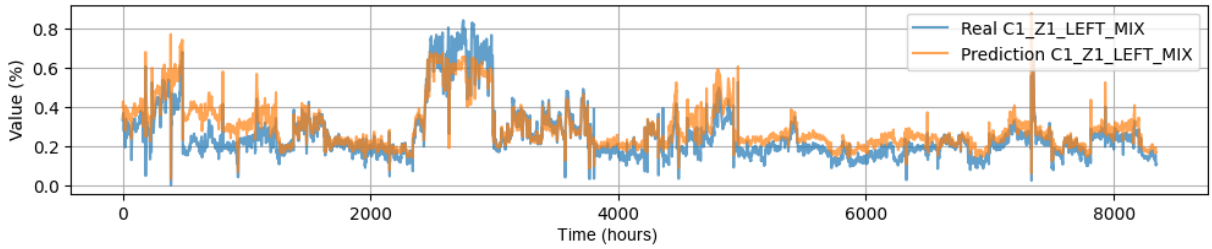

Figure 6: MIX_1 evolution over time in test dataset and M_1MIX prediction. Result $MAE = 0.0529$.

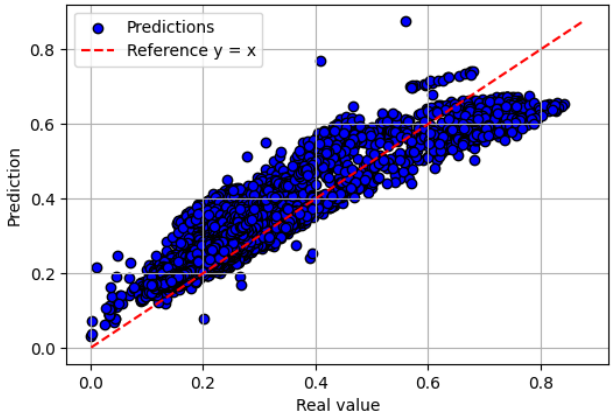

Figure 7: Real values vs predicted values in model M_1MIX. Result $R^2 = 0.76$.

Regarding the explainability of the primary models, it should only be analyzed the coefficient β_j of Equation 1 for each input variable listed in Table 2. Since all primary models share the same linear structure and differ only in the numerical value of their coefficients, the sign and magnitude of β_j provide direct information about the influence of each input variable on the model output. In this regard, it is observed that all estimated coefficients exhibit the *physically expected* sign across the set of primary models, which is consistent with the underlying domain knowledge and confirms the absence of non-physical behaviors learned during training. To further assess the stability of the learned relationships, Table 7 reports the mean value, as well as the minimum and maximum values, of each coefficient β_j computed over all primary models, strengthening the interpretability and robustness of the proposed modeling approach.

However, as shown in Table 6, there is a model that clearly gets a worse result, that is M_3MIX. Despite the accuracy metric, the patterns remain well predicted. Control problems, parameters changes or manual controlling could have happened during the test time range. The nature of the data may be cause of bad accuracy on some cases, but, if carefully trained –by selecting the adequate data– the models are able to capture the correct patterns.

Regarding the surrogate models, their main function is to prevent the optimization algorithm from proposing setpoints that are clearly unrealistic. Consequently, the required accuracy of these surrogate models is moderate, and

Table 7: Statistical summary of the coefficients β_j of the primary models. Models $M_i MIX$ above and models $M_i Cool$ below.

Input variable	Mean β_j	Min β_j	Max β_j
$Side\_Diff\_SP_i$	-0.86	-2.06	-0.33
MIX_{i-1}	0.18	0.03	0.31
$Cool_i$	0.935	0.55	1.42
Pull	-0.18	-0.19	-0.17

Input variable	Mean β_j	Min β_j	Max β_j
$Center\_Diff\_SP_i$	1.08	0.60	1.83
$Cool_{i-1}$	0.75	0.25	1.25
Pull	0.19	0.14	0.25

only significant errors beyond a certain threshold are penalized, to avoid small penalizations that are due to the models inaccuracy. The MAE values of these models are consistently below 10ÂºC for all the test datasets. To show the result of a single surrogate model, in Figure 8 and 9 the predictions of the model $M2_1 Center\_PV$ are plotted.

This requirement is different from the fine temperature stabilization accuracy expected from the industrial control system. A forehearth controller may be required to keep the process temperature within a very narrow band around its setpoint, but the surrogate models proposed here are not used to replace that controller or to close the control loop. Their role is to screen candidate setpoints before they are recommended by the optimizer. For this purpose, the relevant question is whether a candidate setpoint combination is physically attainable and coherent with the observed process behavior, rather than whether the model can reproduce every local temperature fluctuation with a very tight error tolerance. In this way, deviations within the normal uncertainty of the surrogate models are tolerated by design, while large deviations are interpreted as evidence that the optimizer is exploring a setpoint region that the real process is unlikely to reach.

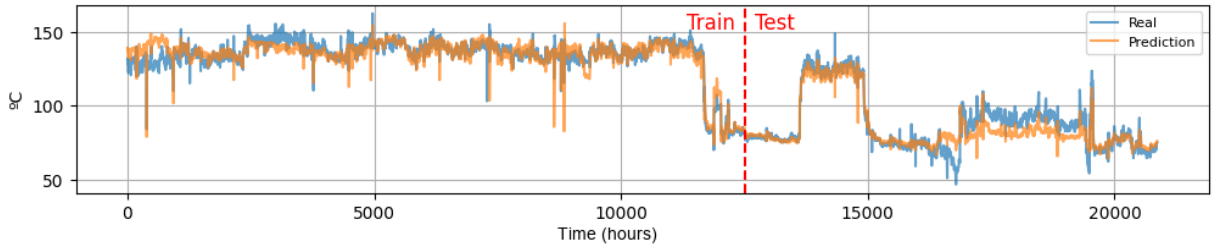

Figure 8: $Center\_Diff\_PV_1$ evolution over time in the whole dataset and $M2_1 Center\_PV$ prediction. Result $MAE = 5.05$.

To enforce this constraint effectively, an exponential penalty approach is implemented through the function g of the Equation 2. Specifically, the Physics Loss (PL) is computed using the absolute error raised to the power of three, as shown in the following equation:

$$PL_i = g(\epsilon_i) = \begin{cases} 3 \cdot \dfrac{\epsilon_i^3}{10^6}, & \text{if } \epsilon_i > 15 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

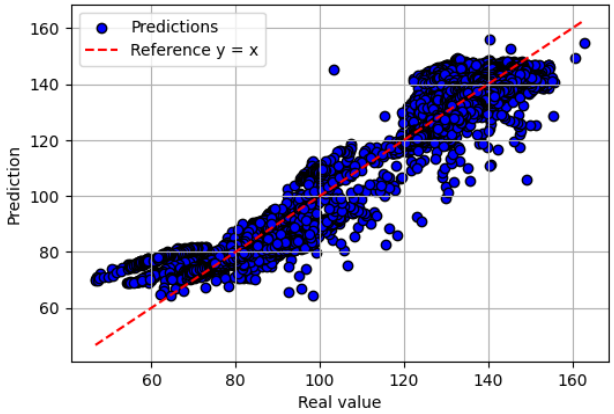

Figure 9: Real values vs predicted values in the whole dataset for model $M2_1 Center\_PV$. Result $R^2 = 0.96$.

where $\epsilon_i = |SP_i - \overline{PV_i}|$. Given that MAE values are consistently below 10ÂºC, any error exceeding 15 Â°C is more likely to reflect a mismatch in the setpoints than a flaw in the model; accordingly, this value was chosen as the threshold. A cubic loss was used to strongly penalize an increase in ϵ_i, as is shown in Figure 10. A different type of function could be used, as the optimization will only be sensitive to the threshold selected; the main purpose of the function is to penalize values above it. The normalization factor 10^6 ensures an appropriate scaling of the penalization term. Once the threshold of 15ÂºC is surpassed, the Equation 7 guarantees a minimum PL value of 0.01. Given that the primary component of the objective function –EC in Equation 5– ranges between 0 and 1 (a percentage, usually below 50%), this minimum penalty equates to a 1% absolute impact on the consumption value, effectively discouraging physically unrealistic solutions. The Equation 7 is the function g of the Equation 2.

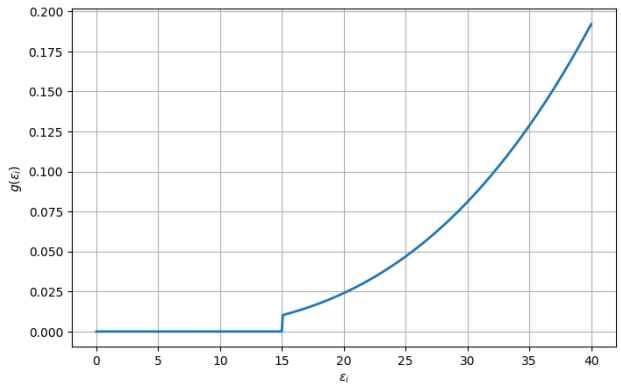

Figure 10: Physic loss value as a function of $\epsilon_i = |SP_i - \overline{PV_i}|$, according to Equation 7.

Regarding the thermal homogenity penalization model, as described in Section 2, it is composed of 9 different linear models, one for each temperature of the 3x3 matrix of temperatures before the spout. However, the interesting value is the final homogenity index, so the evaluation of the models is done based on this value. After a train-test split technique, the obtained MAE over the thermal homogenity index (values within range [80,100]) is 0.88 and an R^2 value of 0.86. An XY scatter plot of the result is shown in Figure 11.

This penalty, named *Homogeneity Penalization*, activates when the homogeneity prediction (hom_{pred}) is below 90. This decision was made based on expertise knowledge, since it is not demonstrated above this threshold that higher

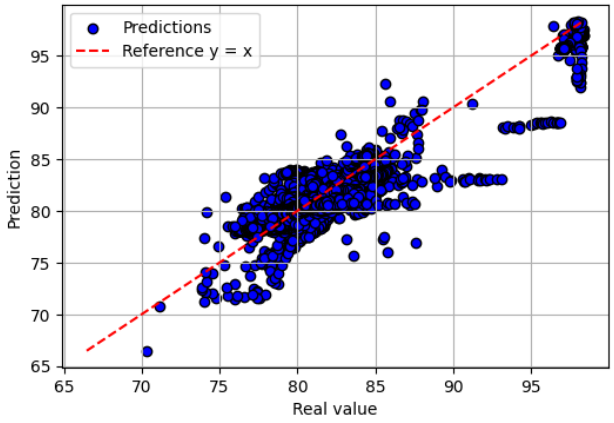

Figure 11: Real values vs predicted values in the homogenity model. Result $R^2 = 0.86$.

homogeneity value means improved process control or better glass distribution. This question is also thoroughly discussed by Groessler[2]. The function h of the Equation 4 to compute the penalty factor is as in Equation 8:

$$P_{hom} = h(hom_{pred}) = \begin{cases} \frac{1}{100}\left(\frac{90 - hom_{pred}}{5}\right)^2, & \text{if } hom_{pred} < 90 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

In this case, a quadratic penalty was selected with the same logic as for the physics loss, to effectively discourage solutions that result in poor homogeneity. As said above, the most important parameter is the threshold selected, and not the type of function, to which the optimization algorithm will not be very sensitive. The *Homogeneity Penalization* sensitivity to the homogeneity value predicted is shown in Figure 12.

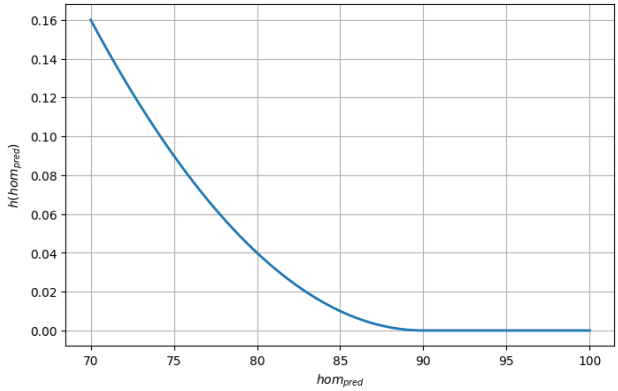

Figure 12: Homogenity loss value as a function of the predicted homogenity value, according to Equation 8.

3.3 Optimization savings

Once evaluated the models trained for use in the objective function, the potential savings of the application of the proposed methodology should be quantified. The result, obviously, can vary significantly depending on the produc-

tion conditions. The method was further tested in a set of input conditions representative of the entire spectrum of possibilitiess for the line (detailed in Table 8), of which some were real and some hypothetical.

Table 8: Input variables range of values for the studied line.

Pull (tones)	WE2 Temperature (Â°C)	Forming Temperature (Â°C)
60-110	1225-1360	1170-1200

For this evaluation, a simple interactive dashboard was designed, where the user inputs the production conditions (Furnace exit temperature, glass pull, glass color and forming temperature) along the actual intermediate setpoints introduced by the operator, that are taken as baseline for comparison. The system then predicts the conditioning outputs (heatings and consumption percentages, coolings, spout temperatures and homogenity) and proposes an optimum set of intermediate setpoints with its corresponding conditioning outputs. The forehearth intermediate setpoints for the evaluated scenarios are always shown in the order: WE1 SP, Z1 Center SP, Z1 Sides SP, Z2 Center SP, Z2 Sides SP. The Table 9 shows the result for 9 different scenarios, covering a wide espectrum of the line working ranges, as stated before. 7 of those cases are real and the last 2 correspond to hypothetical scenarios. Potential absolute savings are always between 1% and 10%, where 0% is the minimum consumption (all burners at minimum power) and 100% the maximum consumption (all burners at maximum power). Relative savings are between 5% and 60%. The average absolute saving for the 9 cases is of 4.91%.

Taking the case 1, the burning and cooling quantities predicted by the models in both cases (actual and optimum proposal) can be compared in Figure 13. It is clear that intermediate setpoints are raised and thus, the cooling outputs (mainly in zone WE1) are reduced. Those changes make the heating quantities to be similar except for the zone 3, where the bigger temperature difference dramatically reduces the heating needed. This makes the total consumption to fall from almost 18.94% to 11.55%, that is a 7.39% of potential absolute saving and 39% of potential relative saving.

Regarding the homogeneity, all the cases fall within acceptable limits in comparison with the baseline case, even it is imporved in some scenarios. The scenario 3 is the one that has the biggest deterioration in homogeneity, with a 5% decrease, but above the lower limit of 90% defined in Equation 8. Additionally, the average difference in homogeneity between baseline and proposed configuration is a slight improvement.

3.4 Operational validation

In addition to the predicted savings discussed above, operational validation was carried out after applying optimized setpoint changes in two real industrial forehearths, both different from the one used as the main case study in this paper. In the first forehearth, the comparison was performed against previous production series of the same production jobs, with comparable pull rate, forming temperature, and furnace exit temperature conditions. Since these data were obtained from normal plant operation rather than from a controlled A/B experiment, the results should be interpreted as industrial validation evidence rather than as a fully isolated experimental campaign. In the second forehearth, the validation was performed as a same-production A/B test: the recommended setpoints were changed during the production campaign, and the optimized period was compared with the same production a few days before the change.

Hourly consumption records were converted directly into energy consumption using the lower heating value of natural gas. The avoided direct CO_2 emissions were estimated by multiplying the energy saving by $0.202\ \mathrm{kg\ CO_2/kWh}$, which is the emission factor assumed for natural gas combustion. This estimate only accounts for direct combustion

Table 9: Optimization results for four scenarios. Production conditions above, with a comparison of baseline versus proposed configurations and outputs below.

	Type	WE2 Temp (Â°C)	Pull (tones)	Color	Forming Temp (Â°C)
1	Real	1358	104	Autumn leaf	1180
2	Real	1225	64.5	Blue	1190
3	Real	1295	105	Flint	1170
4	Real	1328	110	Autumn leaf	1185
5	Real	1274	106	Autumn leaf	1185
6	Real	1285	111	Flint	1185
7	Real	1271	105	Oak	1180
8	Hypothetical	1320	100	Blue	1200
9	Hypothetical	1300	50	Oak	1180

	Configuration	Forehearth Setpoints (Â°C)	Consumption (%)	Homogeneity (%)
1	Baseline	[1270, 1230, 1190, 1205, 1185]	18.94	80.08
	Optimum	[1285, 1236, 1208, 1226, 1205]	11.55 (-7.39)	80.22 (+0.14)
2	Baseline	[1270, 1215, 1230, 1195, 1190]	52.07	92.13
	Optimum	[1256, 1192, 1189, 1182, 1212]	44 (-8.07)	94.37 (+2.24)
3	Baseline	[1270, 1205, 1210, 1180, 1185]	16.8	95
	Optimum	[1263, 1225, 1195, 1217, 1189]	12.7 (-4.1)	90 (-5)
4	Baseline	[1265, 1245, 1230, 1180, 1185]	21.58	83.73
	Optimum	[1277, 1235, 1212, 1206, 1208]	11.92 (-9.66)	82.4 (-1.33)
5	Baseline	[1270, 1230, 1220, 1190, 1190]	22.48	82.92
	Optimum	[1249, 1206, 1191, 1198, 1204]	18.53 (-3.95)	86.39 (+3.47)
6	Baseline	[1270, 1235, 1225, 1190, 1195]	18.88	95.28
	Optimum	[1260, 1217, 1197, 1211, 1202]	15.55 (-3.33)	92.50 (-2.78)
7	Baseline	[1270, 1210, 1190, 1195, 1195]	20.31	81.83
	Optimum	[1262, 1211, 1198, 1190, 1201]	19.19 (-1.12)	84.3 (+2.47)
8	Baseline	[1270, 1230, 1210, 1215, 1215, 1200]	16.57	89.80
	Optimum	[1285, 1235, 1218, 1227, 1221]	15.0 (-1.57)	88.54 (-1.26)
9	Baseline	[1270, 1210, 1190, 1195, 1185]	32.73	80.46
	Optimum	[1263, 1218, 1183, 1199, 1210]	27.68 (-5.05)	83.53 (+3.07)
Average	Baseline		**24.48**	**86.80**
	Optimum		**19.57 (-4.91)**	**86.92 (+0.12)**

emissions.

The operational tests results are displayed in Table 10. All eight operational comparisons show lower measured energy consumption after applying the optimized operating conditions. Depending on the production case and the reference used for comparison, the measured reduction ranges from 26.3 to $89.3\ \mathrm{kWh}$ saved per operating hour, equivalent to relative savings between 8.2% and 22.1%, and 5.3 to $18.0\ \mathrm{kg\ CO_2}$ avoided per operating hour. In the same-production A/B test, represented by case D, the measured consumption decreased from 375.0 to $333.0\ \mathrm{kWh/h}$ after the setpoint change, corresponding to a saving of $42.0\ \mathrm{kWh/h}$, or 11.2%.

For all validation cases, the measured thermal homogeneity remained above 97%. This result is consistent with the fact that the produced glass was flint glass, for which high homogeneity values are generally easier to obtain because its higher heat transfer facilitates thermal equalization across the glass section.

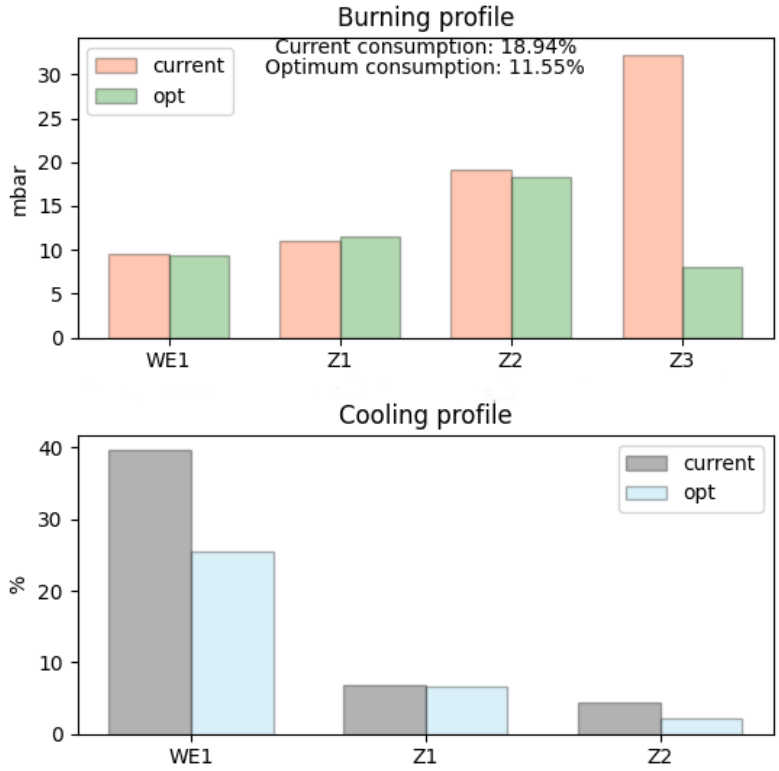

Figure 13: Conditioning output comparison in the first scenario, actual values versus optimum values. Burning air-gas mix pressures above and cooling controller outputs below.

Table 10: Operational validation under industrial production conditions. Cases A–C correspond to three operational tests carried out in one real forehearth; several rows are shown for cases A and C because the optimized operation was compared with more than one previous production series of the same job. Case D corresponds to a second real forehearth, where the validation was performed as a same-production A/B test after changing the recommended setpoints during operation.

Case	**Baseline energy** (kWh/h)	**Optimized energy** (kWh/h)	**Energy saving** (kWh/h)	**Relative saving (%)**	**Avoided CO_2** (kg CO_2/h)
A1	367.5	315.0	52.5	14.3	10.6
A2	404.3	315.0	89.3	22.1	18.0
A3	399.0	315.0	84.0	21.1	17.0
B	414.8	336.0	78.8	19.0	15.9
C1	325.5	294.0	31.5	9.7	6.4
C2	320.3	294.0	26.3	8.2	5.3
C3	351.8	294.0	57.8	16.4	11.7
D	375.0	333.0	42.0	11.2	8.5

4 Conclusions and future work

In this work, a data-driven method to minimize the energy consumption of the glass conditioning process is proposed. The method leverages historical data to train several machine learning models that predict the air-gas mixture output of each forehearth zone. Optimization penalties computed by supervisory surrogate models that predict the intermediate temperatures are applied to prevent overfitting and incorrect model generalizations. A penalization related to the glass thermal homogenity at the outlet of the forehearth is also applied.

The method was tested on a glass container manufacturing line. To evaluate it, the accuracy of trained models and the potential savings in energy consumption were quantified. The prediction models, which predict variables within the range [0,1], obtained a Mean Absolute Error of between $1 \cdot 10^{-2}$ and $13 \cdot 10^{-2}$.

The potential savings of the application of this methodology were found to be highly dependent on the production conditions. Therefore, it was further tested in both standard and extreme situations, of which some were real and some hypothetical. Potential absolute savings were up to 10%, where 0% is the minimum consumption (all burners at minimum power) and 100% the maximum consumption (all burners at maximum power). Relative savings were between 5% and 60%. The average absolute saving for the 9 cases was of 4.91%.

Operational validation was also carried out in two additional real forehearths. Across seven historical comparisons and one same-production A/B test, the measured energy reduction ranged from 26.3 to $89.3\ \mathrm{kWh}$ per operating hour, equivalent to relative savings of 8.2%–22.1% and 5.3–$18.0\ \mathrm{kg\ CO_2}$ avoided per operating hour when only direct natural gas combustion emissions are considered. In all validation cases, the measured thermal homogeneity remained above 97%, supporting that the energy reduction was achieved without compromising the homogeneity constraint.

In addition to energy optimization of the glass conditioning process, this study has also shown the potential of a data-driven system to simplify process modelling and control, providing a viable alternative to the physics-driven approaches. Furthermore, the proposed method can significantly contribute to advancing the decarbonization goals of the glass manufacturing industry.

It must be noted that some criteria used in this work, such as the penalization adjusting functions, are not unequivocal and could be customized depending on the preferences of the conditioning management team. Furthermore, although the forehearth used in the use case is standard within the industry, it can differ from others in terms of geometry, number of zones, cooling techniques, and other factors. For this reason, the methodology will require certain adjustments to suit the particular characteristics of other production lines.

One limitation of this study is the compromised quality of data in certain specific cases. As a result, the trained models exhibit reduced accuracy, although they still meet the initial objective of achieving moderate accuracy and correctly capturing the patterns in glass conditioning.

As future work, hybrid physics-data modelling approaches could be explored to further improve the robustness and generalization capability of the prediction models. In particular, physics-informed neural networks (PINNs) could incorporate simplified heat-transfer constraints or process-consistency terms into the training objective, reducing non-physical extrapolations when the optimizer evaluates operating regions that are weakly represented in the historical data. This could make the setpoint recommendations more reliable across different forehearth geometries, glass colours, pull rates, and thermal operating conditions.

Acknowledgements

Daniel Reguera-Bakhache and Carlos Cernuda are part of the Intelligent Systems for Industrial Systems research group (IT1676-22) supported by the Department of Education, Universities and Research of the Basque Country. The authors would like to thank the Basque Government for the awarding of a grant for an industrial doctorate (Bikaintek 019-B2-2023).

Data Availability

The data supporting the findings of this study are confidential and are not publicly available due to proprietary and confidentiality agreements.

References

[1] Nikolaus Sorg GmbH and Co KG. Glass conditioning, 2013. Accessed: 2025-03-08.

[2] Juergen Groessler. Thermal homogeneity index - the real truth, 2005. Accessed: 2024-06-07.

[3] M. R. Hyre and K. Paul. Effect of natural convection and internal radiation on glass conditioning. *Glass Technology: European Journal of Glass Science and Technology Part A*, 62:160–167, 10 2021.

[4] PSR PARKINSON-SPENCER REFRACTORIES. Glass temperature measurement in forehearths and distributors. *Glass Machinery Plants and Accessories*, 2:66–77, 2014. Accessed: 2025-04-24.

[5] Bill Gough and Don Matovich. Predictive-adaptive temperature control of molten glass. *Record of Workshop Papers - IEEE Industry Applications Society: 1997 Dynamic Modeling Control Applications for Industry Workshop, DMCA 1997*, pages 51–55, 1997.

[6] Alexander Kharitonov and Oliver Sawodny. Flatness-based disturbance decoupling for heat and mass transfer processes with distributed control. *Proceedings of the IEEE International Conference on Control Applications*, pages 674–679, 2006.

[7] Alexander Kharitonov, Sebastian Henkel, and Oliver Sawodny. Two degree of freedom control for a glass feeder. *2007 European Control Conference, ECC 2007*, pages 4079–4086, 2007.

[8] Michal Drapala and Witold Byrski. Continuous-time model predictive control with disturbances compensation for a glass forehearth. *2021 25th International Conference on Methods and Models in Automation and Robotics, MMAR 2021*, pages 366–371, 8 2021.

[9] Michał Drapała and Witold Byrski. Online continuous-time adaptive predictive control of the technological glass conditioning process. *Archives of Control Sciences*, 32:755–782, 2022.

[10] Witold Byrski, Michał Drapała, Jędrzej Byrski, Matti Noack, and Johann Reger. Comparison of lqr with mpc in the adaptive stabilization of a glass conditioning process using soft-sensors for parameter identification and state observation. *Control Engineering Practice*, 146:105884, 5 2024.

[11] Sebastian Henkel, Alexander Kharitonov, and Oliver Sawodny. Modelling and optimisation of a glass feeder considered as a distributed parameter system. *Proceedings of the SICE Annual Conference*, pages 2950–2954, 2007.

[12] Ricky Vesel and Justin Isaacs. Temperature setpoint optimization in steel reheat furnaces using open architecture neural network and modeling software, 2018. Accessed: 2025-03-08.

[13] Zhe Zhao, Yang Du, Yiran Cai, Yu Liu, and Liang Li. Data mining–based modeling of electric energy consumption for electric arc furnace process. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 46(1):1–13, 2024.

[14] Nihan Mirasci and Belgin Aksoy. Forecasting electricity consumption of a cogeneration plant in a food industry company using quantile regression and machine learning algorithms. *Energy Reports*, 11:1780–1790, 2025.

[15] Hamid Pendar, Saman Sadri, Saeid Taherkhani, Mehdi Mozaffar, and Hesam Rafsanjani. Artificial intelligence assisted simulation and optimization of heat and mass transfer in automotive spray painting. *Applied Thermal Engineering*, 229:120331, 2025.

[16] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7:155–162, 1 1964.