A review of the different boiler efficiency calculation and modeling methodologies

Revisión de las diferentes metodologías de cálculo y modelado de eficiencia en calderas

Cristian David Mojica-Cabeza[1]
Carlos Eduardo García-Sánchez[2]
Ramón Silva-Rodríguez[3]
Luis García-Sánchez[4]

[1] Corporación Centro de Desarrollo Tecnológico del Gas (Colombia). Email: cristian.mojica.19@gmail.com
orcid: https://orcid.org/0000-0002-9136-4015

[2] Corporación Centro de Desarrollo Tecnológico del Gas (Colombia). Email: cgarcia@cdtdegas.com
orcid: https://orcid.org/0000-0003-1543-1498

[3] Servicio Nacional de Aprendizaje SENA (Colombia). Email: rsilvar@sena.edu.co

[4] Inline Fluid Systems (Colombia). Email: lgarcia@inlinefs.com

Received: 11-05-2021 Accepted:14-10-2021

How to cite: Mojica-Cabeza, Cristian David; García-Sánchez, Carlos Eduardo; Silva-Rodríguez, Ramón; García-Sánchez, Luis (2021). A review of the different boiler efficiency calculation and modeling methodologies. *Informador Técnico*, *86*(1), 69 - 93.
https://doi.org/10.23850/22565035.3697

Abstract

A review of the different mathematical methodologies for calculating energy efficiency in boilers was carried out in this work, considering both the methods included in standards and the proposals and applications published in research works. The classification was delimited in analytical methods, mechanistic modeling, and empirical modeling; moreover, the main equations for each of the methodologies are presented, which allows building a compilation that is expected to be useful for a first approach to the subject. It is displayed that those mechanistic models are used to evaluate subsystems or specific cases that require a high level of detail, while analytical models are used to make a first approximation to the systems described, and empirical models stand out in terms of their use at the industrial level if there is access to a starting database to adjust them.

Keywords: mathematical modeling; analytical methods; mechanistic modeling; empirical modeling; review; boiler efficiency.

Resumen

En el presente trabajo se realizó una revisión de las diferentes metodologías matemáticas de cálculo de eficiencia energética en calderas, considerando tanto los métodos incluidos en normas como las diferentes propuestas y aplicaciones publicadas en trabajos investigativos. Se delimitó la clasificación en métodos analíticos, modelados mecanicistas y modelados empíricos. Se exponen las principales ecuaciones para cada una de las metodologías, lo que permite construir una compilación, que se espera que sea de utilidad para una primera aproximación a la temática. Se evidencia que los modelos mecanicistas se emplean para evaluar subsistemas o casos puntuales que requieren alto nivel de detalle, mientras que los modelos analíticos se emplean para realizar una primera aproximación a los sistemas descritos, y los modelos empíricos destacan en cuanto al uso a nivel industrial, siempre y cuando se tenga acceso a una base de datos de partida para ajustarlos.

Palabras clave: modelado matemático; métodos analíticos; modelados mecanicistas; modelados empíricos; revisión; eficiencia en calderas.

1. Introduction

Boilers are energy exchange systems that use the heat generated by burning a fuel, transferring it to a stream of water. The resulting stream of hot water or steam is used in industrial processes, heating, or for the generation of electrical energy employing turbines. There are two types of boilers in terms of the distribution of the heat exchange process: pyro-tubular, where the hot flow goes through the tubes, and the water is in the casing, and aqua-tubular, in which the water goes through the tubes and the hot flow through the casing (Kerr; Blair, 2011).

Boilers can use different fuels. Fossil fuels, such as coal, oil derivatives, and natural gas, are the most used (Taler; Dzierwa; Taler; Harchut, 2015), but there are also boilers fueled by industrial waste or biomass, although these present a lower performance than those that use fossil fuels (Dedovic *et al.,* 2012; Kær, 2004). Boilers are very important equipment in the industrial sector and represent a significant part of energy consumption. For example, in Colombia, the most relevant aim uses of energy in industry correspond to indirect heat (e.g., steam) with 44 % of the total energy (Unidad de Planeación Minero Energética [UPME]; Institute for Resource Efficiency and Energy Strategies [IREES]; TEP Energy; Corpoema, 2019).

Energy efficiency is a way of measuring the utilization of the energy available in the fuel used (Chen *et al.,* 2021). Considering the large amount of energy used in steam production in industry, improvements in boiler energy efficiency can lead to large savings. In Colombia, the energy efficiency potential in this sector by implementing the Best Available Technologies and practices worldwide (BAT) increases by 5 to 33 % (UPME *et al.,* 2019). Depending on the measurements available in a boiler, the frequency which both the efficiency and the type of efficiency to be calculated varies, there are several types of mathematical modeling and computational implementations that can be used to calculate efficiency (Amell-Arrieta; Vélez-Rueda, 2003; Valencia-Ochoa; Rojas; Campos-Avella, 2019; Tarasevich; Tepljakov; Petlenkov; Vansovits, 2020; Zhou; Deng; Turner; Claridge; Haberl, 2002).

Typically, the investigation focuses on solving specific problems, and the different reviews found in the literature address those problems. García Sánchez, Chacón-Velasco, Fuentes-Díaz, Jaramillo-Ibarra, and Martínez-Morales (2020) performed an exhaustive review of the state of the art but limited to CDF modelling. Barma *et al.* (2017) describe the implication of boiler's energy efficiency in the environmental impact of the energy sector, and how an improvement in their efficiency would generate considerable reductions in the consumption of fossil fuels and CO_2 emissions (Barma *et al.*, 2017).

In other reviews, Kim, Lee, Tahmasebi, Jeon, and Yu (2021), and Sankar, Santhosh-Kumar, and Balasubramanian (2019) focus on recent methodologies related to computational or numerical simulation. Savargave and Lengare (2018) make a comparison between several Artificial Intelligence (AI) methodologies, but none of them exposes all the existing methodologies. A consolidation from a mathematical standpoint of those methodologies provides a better view of which one of them is most suitable for a given application. The purpose of this paper is to review the different methodologies for determining the energy efficiency of boilers, classify and describe them and mention their suitability for different potential applications.

2. Methodology

The existing information was collected and analyzed between the end of 2020 and the beginning of 2021. Google Scholar was the basic tool for the search, taking advantage of the simultaneous access to searches in different specialized databases, such as Web of Science and Scopus. In addition, some tools were incorporated in specialized databases to broaden the search in an efficient manner, such as the exploration of similar articles, and articles that cite the article viewed. The search equation included a date, must words, and possible words. The date was initially set on 2020 to have access to recent works but eventually was unset to visualize all the existing literature and complete the full picture. Must word in our case were "boiler efficiency" and "review" in the first approach, and then "boiler", "efficiency", and "boiler efficiency". Finally, the search combinations of possible words used were "energy modeling", "mathematical energy modeling", "mathematical modeling", "performance modeling", and "energy performance".

The search combinations made it possible to identify more than 200 articles, which were selected using four levels of filtering to determine the complete analysis of the reference: (i) reading the title, as a first approximation to the content of the article; (ii) determining the year of the reference, prioritizing recent articles; (iii) reviewing the journal where it was published, to discard those that appear in pseudo-scientific journals and other unreliable sources; and (iv) reading the abstract of the article. After the above filtering process, a total of 112 articles were reached; they were read and fully analyzed to perform a final filter, to then select 64 articles that constitute the desired state-of-the-art review.

2.1. Methodologies for the calculation of energy efficiency in boilers

Once the different methodologies for boiler energy efficiency calculation have been reviewed, according to common characteristics, they can be grouped into analytical methods, mechanistic models, and empirical methods (Rusinowski; Stanek, 2007; Savargave; Lengare, 2018). Figure 1 presents the classification of methodologies.

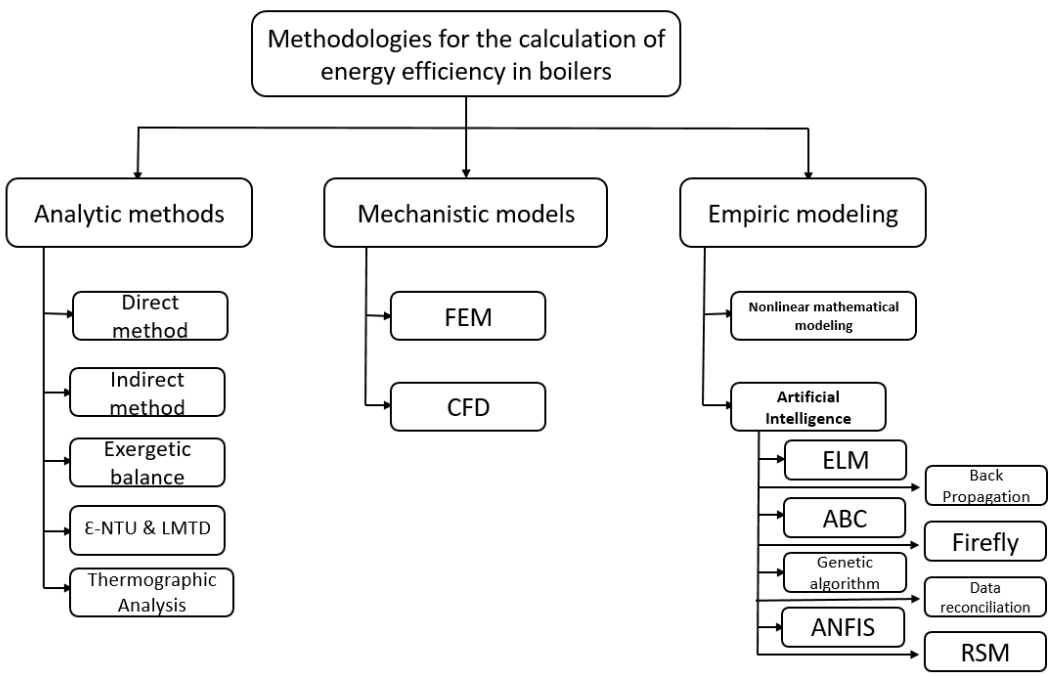

Figure 1. Classification of existing methodologies for calculating energy efficiency.
Source: own elaboration.

Analytical methods

This category includes methods that determine the efficiency of a part of the boiler or the entire one, based on models constructed from energy balances, mass balances, exergy balances, and heat transport equations. The two most important methods in this category are both those that allow the calculation of the so-called direct efficiency and indirect efficiency, which are standardized methods in the Performance Test Code 4 (PTC 4) of The American Society of Mechanical Engineers ([ASME], 2013).

Direct method

The direct method calculates the efficiency by comparing the energy present in the steam output stream and the useful energy present in the fuel used, as shown in Equation 1 (Lang, 2009).

$$\eta_c = \frac{Q_W}{Q_C} * 100 \quad [\%] \qquad (1)$$

In Equation 1, η_c is the direct boiler efficiency, Q_W is the energy in the steam stream, and Q_C is the energy available in the fuel. To determine the useful energy present in the fuel, it is necessary to know its heating value (Kaewboonsong; Kuprianov; Chovichien, 2006). The gross calorific value (GCV) is used, which includes the heat lost by vaporization of the water in the products (Amell-Arrieta; Vélez-Rueda, 2003).

Determining efficiency from the direct method requires accurate and direct measurements of multiple variables. The main variables required are the inlet water flow, the outlet water flow, the secondary outlet flows (blowdowns and auxiliary streams), the pressures and temperatures of the different streams, the fuel flow, and the gross calorific value of the fuel. The calculation of the energy in the steam stream is carried out through Equation 2.

$$Qw = \sum Mr_{S_{z2}} \left(H_{S_{z2}} - H_{E_{z1}}\right) \qquad (2)$$

In Equation 2, $H_{E_{z1}}$ is the specific enthalpy of the fluid entering the system, $H_{S_{z2}}$ is the specific enthalpy of the fluid leaving the system and $Mr_{S_{z2}}$ is the mass flow leaving the system. The energy available in the fuel is calculated according to Equation 3.

$$Qc = MrF * GCV \qquad (3)$$

In Equation 3, *GCV* is the specific gross calorific value, and *MrF* is the fuel mass flow rate (ASME, 2013). *GCV* is calculated according to Equation 4.

$$GCV = \sum_{i=1}^{n} \%(X_i)_{mole} \frac{M_i}{MW_F} GCV_i \qquad (4)$$

M_i, $\%(X_i)_{mole}$ and GCV_i are molar mass, mole fraction, and GCV of the individual components, and MW_F is the molar mass of the fuel.

Indirect method

The indirect method calculates the energy losses in the boiler and subtracts them from efficiency of 100 %. Equations 5 and 6 present the calculation of the indirect efficiency.

$$\eta_c = 100 - \sum q_n \qquad (5$$

$$\sum q_n = q_1 + q_2 + q_3 + q_4 + q_5 + q_6 + \ldots \qquad (6)$$

In Equations 5 and 6, η_c is the indirect boiler efficiency, and q_n corresponds to the different loss terms to be considered in the calculation. For example, the heat loss in the flue gas is calculated according to Equation 7.

$$q_1 = \frac{m_{FG}\, C_{p_{FG}} (T_{FG} - T_a)}{GCV} * 100 \quad [\%] \tag{7}$$

Where m_{FG} is the gas mass flow at the outlet, $C_{p_{FG}}$ is the specific heat of the fuel, and is the ambient temperature.

The number of loss terms can vary according to the level of detail of the modeling. Bujak performed modeling with nine loss terms, including some particularly associated with the use of coal or coal crushing as fuel (Bujak, 2008). Rehan, Habib, Elshafei, and Alzaharnah (2018) used modeling with losses from flue gas, moisture in air and fuel, partial combustion of coal in CO, and radiation and convection. ASME PTC 4.1 uses the following losses: dry flue gas leaving, moisture in the flue gas, moisture in the combustion air, radiation at the boiler surface, and blowdown (ASME, 2013). Heuristic considerations are sometimes used to assign values to some of the loss terms. For example, Qu, Abdelaziz, and Yin (2014) assigned constants 0.015, 0.04, and 0.005 to the last three-loss terms listed previously, based on the 2008 PTC 4.1 (Retirado-Mediaceja *et al.*, 2020). The largest energy losses are considered to occur in combustion, heat exchanger, and flue gas (Barma *et al.*, 2017; Trojan, 2019).

For the indirect efficiency calculation, measurements are required to calculate the loss terms included in the modeling, which implies the need of measuring flue gas temperature, GCV, excess air, thermal properties of flue gas components, temperature, pressure, and ambient humidity (Apaza; Delgado; Garcilazo; Obregón, 2017).

The flue gas temperature is a consequence of energy that ends up heating a non-used stream, although economizers are sometimes used to recover part of this energy. Calculation of the energy lost in the flue gas requires a calculation of the calorific value (C_p) of the stream, which in turn depends on the C_p of the components of the stream. These C_p can be calculated with Equations 8, 9, 10, 11, and 12, where the C_p is given in kJ/kmol*K, for temperatures given in K.

$$C_{p,\,CO_2} = -3.7357 - 4.1034\,\frac{T}{100} + 30.529\left(\frac{T}{100}\right)^{0.5} + 0.024\left(\frac{T}{100}\right)^{2} \tag{8}$$

$$C_{p,\,H_2O} = 143.05 - 183.54\left(\frac{T}{100}\right)^{0.25} + 82.751\left(\frac{T}{100}\right)^{0.5} - 3.6989\left(\frac{T}{100}\right) \tag{9}$$

$$C_{p,\,N_2} = 39.060 - 512.79\left(\frac{T}{100}\right)^{-1.5} + 1072.7\left(\frac{T}{100}\right)^{-2} - 820.4\left(\frac{T}{100}\right)^{-3} \tag{10}$$

$$C_{p,\,O_2} = 37.432 + 0.02\left(\frac{T}{100}\right)^{1.5} - 178.57\left(\frac{T}{100}\right)^{-1.5} + 236.88\left(\frac{T}{100}\right)^{-2} \tag{11}$$

$$C_{p,\,CO} = 69.145 - 0.704\left(\frac{T}{100}\right)^{0.75} - 200.77\left(\frac{T}{100}\right)^{-0.5} + 176.76\left(\frac{T}{100}\right)^{-0.75} \tag{12}$$

The specific heat of the gas corresponds to Equation 13.

$$C_{p_{FG}} = \sum_{i=1}^{k} \% X_{i,mol} C_{pi} \tag{13}$$

The is converted from molar to mass basis using the molecular weight of the gas.

Excess air is the additional amount of air to that stoichiometrically required, which is added to ensure complete combustion. The minimum amount of air is calculated according to the stoichiometry of the combustion reactions, considering that the air contains 21 % oxygen; nevertheless, in practice, complete combustion is not achieved by supplying the minimum air, because the mixture between air and fuel is not perfectly homogeneous, to the low residence time in the chamber, and kinetic issues of the reactions. It may cause incomplete combustion, which results in the generation of carbon monoxide (CO), and negatively influences efficiency. However, excess air also affects energy efficiency, as the incoming air is heated, consuming energy.

For the calculation of the Air-Fuel Ratio (AFR), a generic combustion reaction is shown in Equation (14).

$$\sum_{i=1}^{k}(a_i C_{ni} H_{mi}) + \sum_{i=1}^{k}\left(a_i n_i + \frac{a_i m_i}{4}\right) O_2 + \sigma N_2 \rightarrow \sum_{i=1}^{k}(a_i n_i) CO_2 + \sum_{i=1}^{k}\left(\frac{a_i m_i}{2}\right) H_2O + \sigma N_2 \qquad (14)$$

Where a_i, n_i and m_i are the stoichiometric coefficient and amount of moles of carbon and hydrogen for the i-th species in the fuel, respectively. The theoretically required amount of nitrogen (air) is given by σ in Equation 15:

$$\sigma = \frac{79}{21}\sum_{i=1}^{k}\left(a_i n_i + \frac{a_i m_i}{4}\right) + \sigma_F \qquad (15)$$

To calculate the excess air required, the term βO_2 is added in products for the combustion reaction, so that the moles of nitrogen, σ, change to Equation 16.

$$\sigma = \frac{79}{21}\left(\sum_{i=1}^{k}\left(a_i n_i + \frac{a_i m_i}{4}\right) + \beta\right)\sigma_F \qquad (16)$$

Exergetic balance

Exergy is the maximum amount of work available from a flow and is calculated by bringing the flow to a thermodynamic equilibrium state with reference (ambient) conditions. An exergy balance simultaneously assesses the quantity and quality of energy associated with the process (Behbahaninia; Ramezani; Lotfi Hejrandoost, 2017). It can also be interpreted as a measure of the energy irreversibilities associated with this process since unavoidable energy losses are not energetically equivalent (Lozano; Valero, 1993).

Behbahaninia *et al*. (2017) performed a parametric analysis of destruction and efficiency concerning reference conditions T = 25 °C and P = 1 bar (Lang, 2009), they found that for an increase in temperature, the destruction increases and the efficiency decreases. The calculation associated with exergy can be divided into subsystems: exergy destruction in the boiler, convection losses, destruction in the heater, loss in the emitted gas, loss due to CO formation, and loss due to unburned fuel.

Like energy efficiency, the greatest exergy loss occurs at the burner, followed by the heat exchanger. The blowdown is not considered a loss, but a product of exergy as it depends on the quality of the water and not on the efficiency of the boiler. Inside the air mixer, there is no energy loss, but there is an exergy loss associated with mixing and heat transfer. Kinetic and potential energy are not considered in these balances. Losses due to radiation and incomplete combustion are negligible for a properly functioning system but should be regarded if the burner or insulation is considered to warrant it. Briefly, the balance could be reduced to product exergy, fuel exergy, losses, and destruction (Behbahaninia *et al.*, 2017; Dorotić; Pukšec; Duić, 2020; Farhat; Zoughaib; El Khoury, 2015).

To obtain the exergy losses, mass, energy, and exergy balances must be established, leading to Equation 17.

$$\dot{E}_F = \dot{E}_p + (\dot{E}_{L1} + \dot{E}_{L2} + \dot{E}_{L3} + \dot{E}_{L4}) + (\dot{E}_{D1} + \dot{E}_{D2}) \qquad (17)$$

$\dot{E}_F$ is fuel exergy, $\dot{E}_p$ is exergy in products, $\dot{E}_{L,i}$ terms are exergy losses and $\dot{E}_{D,j}$ terms represent exergy destruction. $\dot{E}_F$ contains 3 components, as shown in Equation 18.

$$\dot{E}_F = \dot{E}_f + \dot{E}_{AS} + \dot{E}_{a,11} \qquad (18)$$

$\dot{E}_f$ is the chemical exergy of the fuel consumed, $\dot{E}_{AS}$ is the exergy of the atomized stream, and $E_{a,11}$ is the physical exergy of the air.

$\dot{E}_p$ can be calculated according to Equation 19, where $\dot{m}_i$ are the mass fluxes of the different components and ε_i are the exergies of the components.

$$\dot{E}_P = \sum \dot{m}_i{}^*\varepsilon_i \qquad (19)$$

Among the exergy losses, $\dot{E}_{L1}$ is that associated with the stack gas, according to Equation 20.

$$\dot{E}_{L1} = \frac{\dot{m}_G}{M_G}{}^*\tilde{\varepsilon}_{ch} + \dot{m}_G{}^*\tilde{\varepsilon}_{ph} \qquad (20)$$

M_G is the equivalent molecular mass of the gas, product of the molar mass of the components multiplied by their mole fraction, and is the mass of gas measured at the stack exit. $\dot{E}_{L2}$ is the exergy associated with the unburned fuel, which for the case of gaseous fuels is assumed to be zero. $\dot{E}_{L3}$ is the exergy associated with the incomplete combustion emission and CO formation.

For the case of $\dot{E}_{L4}$, associated with exergy dissipation through the boiler surface, it is calculated with Equation 21 (ASME, 2013).

$$\dot{E}_{L4} = Q_{radiation}{}^*\left(1-\frac{T_0}{T_s}\right) \qquad (21)$$

Where T_0 is the ambient temperature and T_S is the boiler surface temperature.

As for the exergy destruction terms, two aspects are considered: one is associated with the air heater, which is not considered for all types of boilers. The other corresponds to the exergy destroyed in the boiler, according to Equation 22 (Behbahaninia *et al.*, 2017).

$$\dot{E}_{D1} = \dot{E}_F - \dot{E}_p - \left(\frac{\dot{m}_G}{M_G}{}^*\tilde{\varepsilon}_{ch} + \dot{m}_G{}^*\tilde{\varepsilon}_{ph}\right) - \dot{E}_{L4} \qquad (22)$$

A direct exergy efficiency can be calculated using the exergy losses according to Equation 23.

$$\eta = \frac{E_p}{E_F} \qquad (23)$$

Alternatively, the indirect exergy efficiency can be calculated with Equation 24.

$$\eta = 1 - \frac{\text{destroyed exergy}}{\text{incoming exergy}} = 1 - \frac{\Sigma(\dot{E}_L\dot{E}_D)}{\dot{E}_F} \qquad (24)$$

LMTD and Ɛ-NTU

The Logarithmic Mean Temperature Difference (LMTD) is used to estimate the temperature associated with heat transfer in heat exchangers. Thus, this approach can be used to calculate the efficiency of the heat exchange subsystem in a boiler. The LMTD is a logarithmic approximation of the temperature difference between the heat exchanger inlets and outlets; its value is directly proportional to the heat transferred and can be calculated from Equation 25.

$$LMTD = \frac{\Delta T_A - \Delta T_B}{ln\left(\frac{\Delta T_A}{\Delta T_B}\right)} = \frac{\Delta T_A - \Delta T_B}{ln\Delta T_A - ln\Delta T_B} \qquad (25)$$

Where ΔT_A is the temperature difference between the two streams at point A and ΔT_B is the temperature difference at point B.

The number of units transferred (NTU) method is an alternative for estimating the heat transfer rate in heat exchangers when not all the data needed to do so through LMTD are known. The NTU method can be applied regardless of the flow distribution, either parallel or counterflow (Trojan, 2019), according to Equations 26, 27, and 28.

$$\varepsilon = \frac{1 - \exp\left(-NTU^*\left(1 - \frac{\dot{C}_{min}}{\dot{C}_{max}}\right)\right)}{1 - \frac{\dot{C}_{min}}{\dot{C}_{max}} * \exp\left(-NTU^*\left(1 - \frac{\dot{C}_{min}}{\dot{C}_{max}}\right)\right)} \tag{26}$$

$$NTU = \frac{k^*A}{\dot{C}_{min}} \tag{27}$$

$$\dot{C}_{min} = min\left[\left(C_p\dot{m}\right)_{gas}, \left(C_p\dot{m}\right)_{steam}\right] \tag{28}$$

Where $\dot{C}_{min}$ is the minor between steam and flue gas flow, k is the transfer constant and A is the heat exchange surface area (Modliński; Szczepanek; Nabagło; Madejski; Modliński, 2019).

Thermographic analysis

The thermographic analysis is e to establish radiation losses, using a thermal camera to record boiler surface temperatures. This allows estimating heat transfer losses in the walls of the boiler, radiation losses in different areas of the boiler, and detecting insulation problems. This analysis is important insofar as radiation losses are energy that is directly dissipated to the environment, diminishing the efficiency of the boiler.

The way to perform the analysis is based on the sum of the surface areas where the temperature is above the expected. The analysis also can be carried out by determining what percentage of the surface area has such unacceptable or worrying temperatures. These situations can be caused by poor burner adjustment or poor insulation. The interpretation of the histogram is done through areas; the total area with a certain temperature is directly proportional to the histogram height for the same temperature (Jiménez-Borges; Madrigal; Lapido; Vidal, 2016; Serway; Jewett, 2008). Figure 2 illustrates the process of loss analysis by thermography.

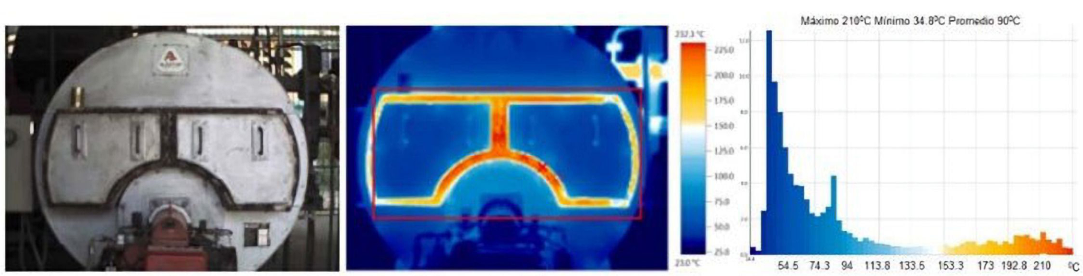

Figure 2. Boiler front surface, thermographic camera shot, and histogram with the corresponding distribution. Source: Jiménez-Borges *et al.* (2016).

2.2. Mechanistic Models

Mechanistic models divide a complex system into subsystems or parts. To understand the behavior of the components of a system, mechanistic models assume that a system can be understood by analyzing how the different parts perform together and separately. Typically, mechanistic models are associated with a physical, tangible system whose components are solid and visible.

FEM

The Finite Element Method, FEM, is a numerical method employed in the solution of partial differential equations associated with engineering and physics problems. It is used in the design of industrial applications and the simulation of complex physical systems (Tognoli; Najafi; Rinaldi, 2018). The development of a FEM algorithm to solve a problem requires a four-stage development: (1) formulation of the problem in a variational form, (2) division of the domain of independent variables into finite elements, (3) projection of the original variational problem onto the vector space and (4) numerical computation of the solution of the system of equations. These steps allow representing a differential calculus problem with a linear algebra problem. The problem is posed on a vector space of non-finite dimension; nevertheless, it can be solved approximately by finding a projection onto a subspace of finite dimension (Zhang; Yang; Hu; Lu; Wu, 2013).

To address the FEM methodology, a case study presented by Tognoli *et al*. (2018) will be discussed, in which the boiler is mainly divided into the gas side and water/steam side. For the steam side, the optimal value of NF was determined to be 100 and of Nj to be 20, considering the relationship between error and computational cost. The distribution of the partitioning system can be seen in Figure 3.

For each small system, a mass balance is performed, corresponding to inputs, outputs, consumption, and generation. Under the assumption that the surfaces behave as gray bodies, the Dittus-Boelter correlation is used for the convective part. Correlations are considered to model the heat exchange between the flue gases in the hearth and the water in the casing. To simulate the water-steam system (casing), three main flows must be considered: $\dot{m}_f$ is the water feed flow, is the steam outflow and $\dot{m}_p$ is the blowdown outflow. These assumptions generate several equations, which are used in conjunction with the main energy balance shown in Equation 29.

$$\frac{d}{dt}T \cong \frac{1}{\rho V_{total}\underline{C}_V}\left[\dot{m}_f\underline{C}_{pf}\left(T_f-T\right) - \dot{m}_v^{+}\underline{\lambda} + \dot{q}_{F,water} + \sum_{j=2}^{3,4}\dot{q}_{j,water}\right] \quad (29)$$

Consequently, the main efficiency of the boiler is calculated according to Equation 30.

$$\eta_{boiler,\ l} = \frac{\sum_{j=1}^{j=Ms}\left(m_{vj}^{+}h_{vj}^{+} - m_{fj}h_{fj}\right)}{\sum_{j=1}^{j=Ms}\left(m_{fuelj}{*}PCI\right)} \quad (30)$$

Where *j* corresponds to the simulation time interval, from 1 to the end of the simulation. The length of the time intervals remains constant (dt).

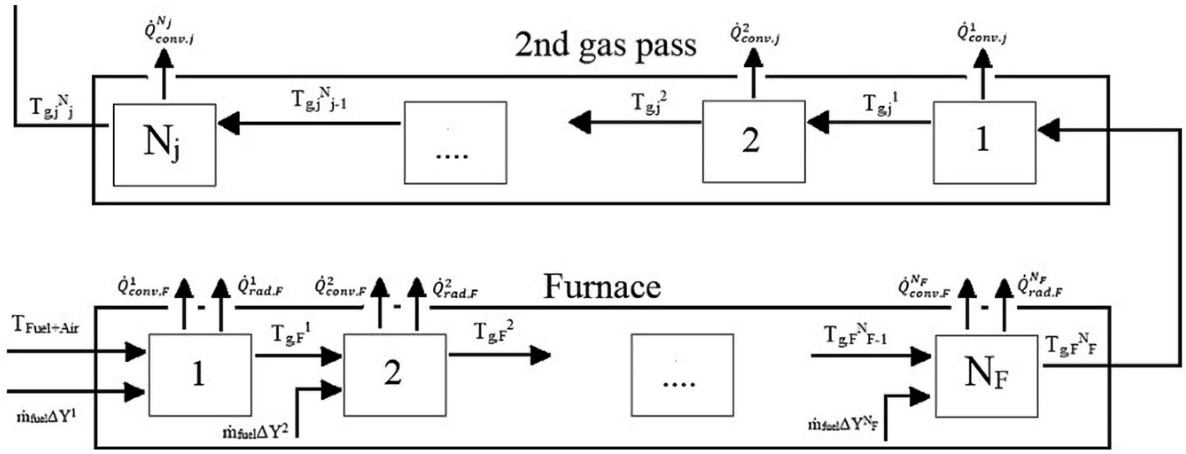

Figure 3. Structure of a FEM system for a boiler.
Source: Tognoli *et al.* (2018).

CFD

Computational Fluid Dynamic Modeling (CFD) is a complex methodology that models in detail the behavior of liquids and gases from their physicochemical properties, generating a three-dimensional model of the fluids, as shown in Figure 4 (Hasini; Yusoff; Shuaib; Boosroh; Haniff, 2009). The physicochemical properties are deeply linked to the temperature of the streams, so heat transfer, either by convection or radiation, is successfully simulated. However, it requires vast computational power to generate such complex systems, so their applicability is often limited to case studies or academia (Díez; Cortés; Campo, 2005; Tognoli *et al.*, 2018).

Accordingly, when the level of detail required warrants a CFD simulation and there is no need for real-time calculations, CFD can be applied for a specific subsystem. Among the subsystems typically modeled with this methodology are the heat exchanger and the convective heat transfer zone (Milićević; Belošević; Crnomarković; Tomanović; Tucaković, 2020; Modliński *et al.*, 2019). The detailed CFD calculation is based on simulations of previously designed programs such as CFX, CFD ACE+, or CFD GEOM (to generate the spatial distribution mainly).

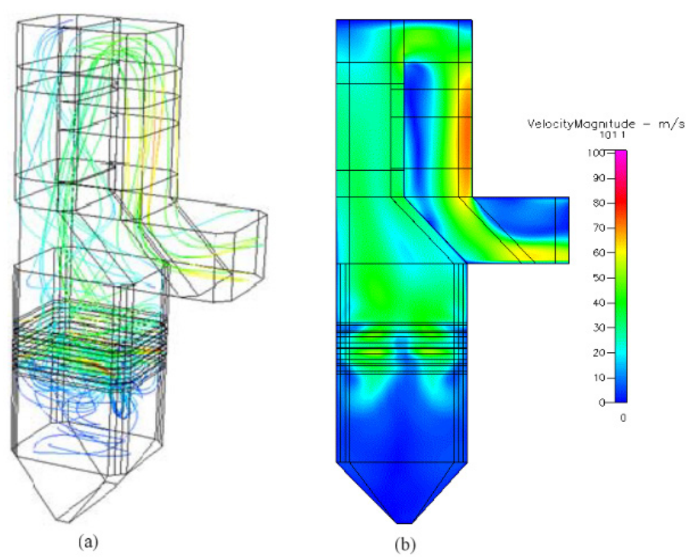

Figure 4. Example of CFD simulation. (a) Flow pattern. (b) Velocity distribution.
Source: Hasini *et al.* (2009).

The success is to show 3D modelings of the different systems (Abroshan, 2020). So the geometric dimensions of what is to be simulated, the flow direction, and the different implications of the process as such are chosen. Some of the parameters such as gas composition, velocity, temperature profiles for the different flows, mass balances, energy balances, and the different assumptions are taken as an input when modeling the system, which is provided directly to the software (Gutiérrez-Ortiz, 2011; Hasini *et al.*, 2009; Math *et al.*, 2021).

2.3. Empiric methods

Empirical models are not characterized by a specific type of mathematical expression but are based on fitting various modeling approaches to available data (Qin; Li, 2020). This category includes models whose mathematical presentation depends on the input and output variables of interest chosen and the modeling approach selected.

Non-lineal mathematical modeling

This category includes modeling approaches in which a non-linear model is fitted. It aims to find the most influential variables in the system, the output variables of interest, and generate a model that relates them (Abubakar; Bello; Ejilah, 2020; Ivanitckii; Sultanov; Kuryanova, 2021; Zhitarenko; Bejan; Ostapenko, 2020). For example, in one study, from a 3E (Efficiency, Economy, Environment) projection, basic energy calculations are handled, and fuel energy availability is calculated (Zhao; Duan; Liu, 2019). On the other hand, Rusinowski and Stanek (2007) describe a methodology that, based on the DIN 1942 standard, raises the material and energy balances, thus generating a database that is subsequently used as input for a neural network. Bujak (2008) defines which terms of the energy balance must be included in the mathematical model due to their influence on the efficiency so that the efficiency can be estimated with fewer variables.

Multiple Linear Regression

This type of regression is used to relate two or more independent variables, called regressors, to a dependent variable, as shown in Equation 31.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i \qquad (31)$$

Where the regression coefficients β_k denote the magnitude of the effect that the regressors X_{ki} has on the independent variable Y_i, β_o is the independent term of the model and ϵ_i is the random error term of the model. The multiple linear regression model can also be represented in matrix form, according to Equations 32, 33, 34, 35, and 36.

$$Y_i = X\beta + \epsilon \qquad (32)$$

$$y = [y_1\, y_2 \vdots\, y_n\,] \qquad (33)$$

$$X = [1\; x_{11}\, x_{12}\; 1\; x_{21}\, x_{22} \vdots 1\; \vdots x_{n1}\; \vdots x_{n2}\; \ldots \ldots \ddots \ldots\; x_{1k}\, x_{2k} \vdots x_{nk}\;] \qquad (34)$$

$$\beta = [\beta_0\, \beta_1 \vdots\, \beta_k\,] \qquad (35)$$

$$\epsilon = [\epsilon_1\, \epsilon_2 \vdots\, \epsilon_n\,] \qquad (36)$$

The model parameters can be fitted using several methods. The most popular is the least-squares one, in which the set of parameter values is chosen that minimizes the sum of the squared differences between the values estimated by the model and the experimental data (Granados, 2016).

Principal Component Analysis

Principal Component Analysis (PCA) is a descriptive statistical technique that allows obtaining a model with a smaller number of variables (characteristics or regressors), trying not to lose information from the original model. Suppose that there is a sample with n individuals with m variables F_j measured for each one. Using PCA, several factors $p < m$ are sought that approximately explain the value of the m variables for everyone. The method to apply the PCA starts from the correlation matrix, considering the value F_j of each of the m random variables. As shown in Equation 37, the value of these variables is written for each of the individuals in the form of a matrix.

$$\left(F_j^{\beta}\right)_{j=1,\ \dots,\ m}^{\beta=1,\ \dots,\ n} \tag{37}$$

Equation 38 presents the sets M_j that can be considered random samples for F_j.

$$M_j = \left\{F_j^{\beta} \mid \beta = 1,\ \dots,\ n\right\} \tag{38}$$

From the $m \times n$ data corresponding to the m random variables, a sample correlation matrix is constructed, defined as shown in Equations 39 and 40.

$$\mathrm{R} = \left[r_{ij}\right] \in M_{mxm} \tag{39}$$

$$r_{ij} = \frac{cov\left(F_i, F_j\right)}{\sqrt{var(F_i)\ var(F_j)}} \tag{40}$$

Since the correlation matrix is symmetric then it is diagonalizable, and its eigenvalues satisfy Equation 41.

$$\sum_{i=1}^{m} \lambda_i = m$$

Due to the above property, these m eigenvalues are called the weights of each of the m principal components. The mathematically identified principal factors are represented by the eigenvector basis of the matrix R. Then each of the variables can be expressed as a linear combination of the eigenvectors or principal components (Forkman; Josse; Piepho, 2019). Bahadori and Vuthaluru (2010) use PCA to set a model that defines the boiler's energy efficiency concerning excess air at the combustion.

Artificial Intelligence

In this section, the different methodologies associated with artificial intelligence are grouped. In all cases in this category, the model output variable is a measure of boiler energy efficiency. The application of artificial intelligence techniques seeks to determine efficiency using different or partial information, regarding the measurements required for an analytical calculation of efficiency. Some of the methodologies combine artificial neural networks with evolutionary computing algorithms to determine the model parameters. These algorithms are based on imitating nature behaviors, such as bee colony, firefly algorithm, and genetic algorithms. The main application in the calculation of energy efficiency in boilers is modeling from databases, which allows the generation of equations or algorithms that, with a few input variables, can estimate efficiency quite accurately (Tang; Li; Kusiak, 2020).

Artificial Neural Networks

Artificial Neural Networks (ANN) are modeling methods inspired by the functioning of neurons. This system is divided into layers: an input layer and one or more intermediate hidden layers that generate the variables of interest (output), as shown in Figure 5.

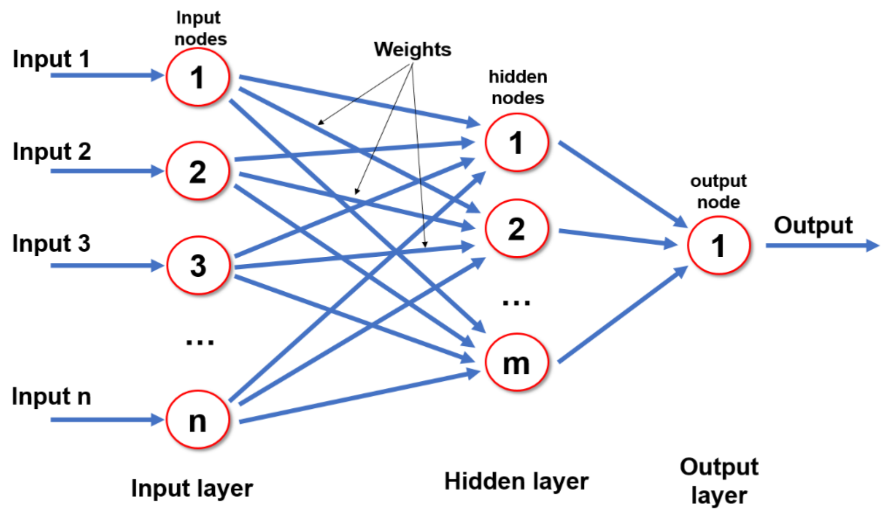

Figure 5. The basic structure of an ANN.
Source: own elaboration.

The connection between neurons in each of the layers is mediated by a parameter Θ or weight, a value obtained through training (fitting to data) of the network (Ding; Liu; Xiong; Jiang; Shi, 2018; Irwin; Brown; Hogg; Swidenbank, 1995; Rusinowski; Stanek, 2007). Once the network training (generation of coefficients) from a database is done, the network is considered ready and can be applied to generate predictions (Saha; Shoib; Kamruzzaman, 1998).

In terms of notation, $a_i^{(j)}$ can be understood as the activation of unit i in layer j of the network, and $\theta^{(j)}$ as the matrix of weights controlling the mapping of the function from layer j to layer j +1. Thus, a system with 3 input variables, a hidden layer with 3 units, and an output variable would be represented by Equations 42 , 43, 44, and 45.

$$a_1^{(2)} = g\left(\theta_{10}^{(1)}x_0 + \theta_{11}^{(1)}x_1 + \theta_{12}^{(1)}x_2 + \theta_{13}^{(1)}x_3\right) \quad (42)$$

$$a_2^{(2)} = g\left(\theta_{20}^{(1)}x_0 + \theta_{21}^{(1)}x_1 + \theta_{22}^{(1)}x_2 + \theta_{23}^{(1)}x_3\right) \quad (43)$$

$$a_3^{(2)} = g\left(\theta_{30}^{(1)}x_0 + \theta_{31}^{(1)}x_1 + \theta_{32}^{(1)}x_2 + \theta_{33}^{(1)}x_3\right) \quad (44)$$

$$h_\theta(x) = a_1^{(3)} = g\left(\theta_{10}^{(2)}a_0^{(2)} + \theta_{11}^{(2)}a_1^{(2)} + \theta_{12}^{(2)}a_2^{(2)} + \theta_{13}^{(2)}a_3^{(2)}\right) \quad (45)$$

Where h_θ is the final output of the neural network, i.e., the efficiency.

Usually, multiple training iterations are required to find the neuron weight that gives the smallest difference between the desired value, z_j, and the network output, the main objective being to reduce the quadratic error of the response, as shown in Equation 46.

$$\min E = \frac{1}{2} \sum_{j=1}^{N} (z_j - y_j)^2 \quad (46)$$

The selection of the network topology (number of layers and number of elements in each layer) is not a problem with a unique solution. A suitable topology depends on the complexity of the problem and the size of the training set. In general, it is advisable to start with a simple network and increase the degree of complexity (Lawrence; Giles; Tsoi, 1996). Too few nodes will lead to a high error for the system as the predictive factors may be too complex for a small number of nodes to predict. On the other hand, too many nodes will adapt too much to the training set, presenting problems of overfitting, i.e., the network will be useless for data moderately different from the training data (Bengio; LeCun, 2019). Maddah, Sadeghzadeh, Ahmadi, Kumar, and Shamshirband (2019) model efficiency as a function of temperature steam and flow rate of the generated steam with 93 inputs, using 70-15-15 distribution to training, validation, and test. The structure of the resulting ANN is 2-5-1 with an error of 0.8 %.

ELM

Extreme Learning Machine, ELM, is a hidden-layer feed-forward ANN in which no initial values are needed, since the input weights and biases are generated randomly, increasing the randomness of the system concerning others in which initial values are fixed. It has a fast learning algorithm and good generalization capability and easily overcomes problems such as local minimum and stopping criteria. Suppose there are N samples (x_i, t_i), where $x_i = [x^i_1, x^i_2, \ldots, x^i_n]$ is an n-dimensional vector of the i-th sample and $t_i = [t^i_1, t^i_2, \ldots, t^i_L]$ is the target vector. Having W as an input weight of dimensions $M \times n$, B as hidden layer bias of dimensions $M \times 1$, and β as the output weight of dimensions $L \times M$. The output (T) of the ELM with M hidden neurons can be calculated according to Equations 47, 48, 49, and 50.

$$H\beta = T \tag{47}$$

$$\mathrm{H}(W,B,X) = \begin{bmatrix} g(w_1x_1+b_1) & \cdots & (w_Mx_1+b_M) \\ \vdots & \ddots & \vdots \\ g(w_1x_N+b_1) & \cdots & (w_Mx_N+b_M) \end{bmatrix}_{NxM} \tag{48}$$

$$\beta = \left[\beta_1 \ldots \beta_M\right]^T_{LxM} \tag{49}$$

$$T = \left[t_1 \ldots t_N\right]^T_{LxN} \tag{50}$$

Then the output weight β is determined by Equation 51 analytically.

$$\tilde{\beta} = arg \min_{\beta} \|H\beta\text{-}T\| = H^{+}T \tag{51}$$

Where H^+ is the generalized Moore-Penrose inverse of H. If the condition of $rand(H) = M$ is satisfied, Equation 51 can be rewritten as Equation 52.

$$\tilde{\beta} = \left(H^TH\right)^{-1}H^TT \tag{52}$$

Li, Niu, Liu, and Zhang (2012) use ELM to obtain an empirical relation between combustion efficiency and operational variables of boilers.

Artificial Bee Colony (ABC) Algorithm

The Artificial Bee Colony (ABC*) a*lgorithm is used to optimize the input weights and biases of the hidden layers. It is based on the behavior of three classes of bees: worker, spectator, and scout bees. Each worker bee is associated with a single food, which implies that the number of worker bees is equal to the number of food sources. Workers make a journey to the food and return to the hive, when they no longer find the food, they become scouts and must search for a new one.

The exploration process is related to the ability to independently search for a global optimum, while the operation process is related to the ability to apply existing knowledge to search for better solutions. This algorithm was employed to optimize the ELM model (Li; Niu; Liu *et al.*, 2012).

Back Propagation

The Back Propagation (BP) algorithm is the most widely used for training ANN. The main advantage of BP is that it considers all the weights for each layer, avoiding redundant computations that could arise in intermediate terms for networks with some complexity in their topology (Kljajić; Gvozdenac; Vukmirović, 2012). To apply BP, the delta rule of BP can be followed, in which the values of differences zc($\delta = z_j - y_j$) are determined based on the values of the next layer and the weights in connection with the hidden layer and the next layer. The BP process starts with the calculation of δ for the output layer and then going backward, the errors are propagated throughout the entire neural network. The problem of minimizing the objective function can be solved by the gradient method described by Equation 53.

$$w(s+1) = w(s) - \eta \nabla E\big(w(s)\big) + \alpha \Delta w\ (s - 1) \qquad (53)$$

This equation is applied in the training process to determine the values of the neuron weights. The learning process begins with the successive introduction of operational points from the learning set into the inputs of the neural network. Then, the delta values (errors) are calculated for the output layer, and the calculated errors are propagated backward through the network, and finally, the weights are corrected. This sequence is repeated for all points in the training set. After this process the entire network weight matrix is determined, leaving the system ready for simulation. If the error is less than expected the system is ready, otherwise the process is repeated from training (Rusinowski; Stanek, 2010).

Firefly Algorithm

The Firefly (FF) Algorithm is used to train and optimize an existent ANN based on the idea that fireflies are attracted based on the intensity of brightness; thus, an initial input value is assigned, and a specific brightness is set to the objective function (Savargave; Lengare, 2018). It needs input and output values for training the ANN. It presents higher accuracy than BP for nonlinear correlations (Savargave; Lengare, 2017).

The FF follows two main rules: (1) the landscape of the objective function evaluates the brightness of fireflies, and (2) the brightness and attraction of fireflies are proportional to each other, and both decrease with increasing distance. Since the attraction of the fireflies is directly proportional to the intensity of the light emitted by the other firefly, the changes in attraction β concerning distance r can be calculated by Equation 54.

$$\beta = \beta_0 e^{\gamma r^2} \qquad (54)$$

Where β_0 represents the attraction when $r = 0$. The motion of firefly i toward the glow shown by firefly j can be evaluated with Equation 55.

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j^t - x_i^t\right) + \alpha_t \epsilon_i^t \qquad (55)$$

Where the second term is formed due to the attraction between fireflies, and the third term is a random motion with $\alpha_t \epsilon_i^t$ representing the number selected in a random way using the uniform Gaussian distribution for some time t, y α_t is the randomization parameter. When $\beta_0 = 0$, the firefly chooses the random motion, and if $\gamma = 0$ a minimum is obtained (Savargave; Lengare, 2017).

Genetic algorithm

The Genetic algorithm (GA) is based on the process of natural selection, emulating methods of nature such as mutation or crossover of genotypes, thus searching for solutions to certain problems. It is often used to optimize problems containing function-free models that cannot be optimized by normal methods. Zhang, Ding, Wu, Kong, and Chou (2007) use GA to optimize an ANN of NO_x emission and efficiency.

The first step in the implementation of any genetic algorithm is the generation of the initial population. Each member of the population will be a binary string of length L corresponding to the problem encoding, which mimics a genotype or chromosome. Then, each string is evaluated according to the evaluation function, or objective function, which dictates the performance concerning a specific set of parameters. According to the value of the objective function, reproductive opportunities are assigned, whereby a selection process occurs that mimics that which occurs in sexually reproducing populations of living beings. The process of generational change is illustrated in Figure 6.

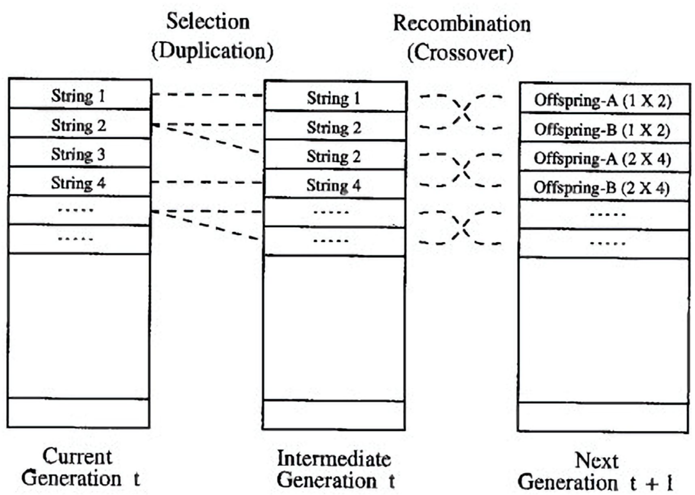

Figure 6. Generational change in the genetic algorithm.
Source: Whitley (1994).

After recombining, the mutation operator can be applied. For each bit in a string, the mutation occurs with a probability of less than 1 %. In other words, 1 bit of a string is changed for every 100 existing bits in case of a probability of 1 %. Once the selection, recombination, and mutation process are completed, the new population can be evaluated. The process of evaluation, selection, recombination, and mutation forms a generation in the execution of a genetic algorithm (Whitley, 1994).

Data Reconciliation

It consists of taking data that present a certain error, seeking to reduce it. Reconciliation is usually used to correct errors in data obtained when making energy and mass balances, which normally are used as input of the empirical methods and AI applications (Szega, 2020). Errors in the measurement results may be due to the inaccuracy of the equipment, failures, or poor signal processing. Data reconciliation allows to have higher reliability in the measurements of process variables, to evaluate the accuracy of the adjusted results of the measurements, and to decrease the uncertainty of the measurements taken (Szega; Nowak, 2015). The data reconciliation problem for steady-state processes represents an optimization problem with constraints, as shown in Equation 56.

$$\sum_{i=1}^{N_{med}} \rho(\varepsilon_i)$$

Subject to:

$$\begin{aligned} & f_k(x, z) = 0 \quad k \in N \\ & g_l(x, z) \geq 0 \quad l \in N_{des} \\ & x_i^{inf} \leq x_i \leq x_i^{sup} \quad i \in N_{med} \\ & z_j^{inf} \leq z_j \leq z_j^{sup} \quad j \in N_{unmed} \end{aligned} \quad (56)$$

N_{med} is the number of measurable variables; N is the number of equality constraints, in this case, the number of equations; N_{des} is the number of inequality constraints; N_{unmed} is the number of unmeasurable variables; $\rho(\varepsilon_i)$ is the objective function; f is the equality constraint, in this case, the mass or energy balance; g is the inequality constraint imposed on the problem; and z is the unmeasured variables of the process, estimated with reconciliation.

In the objective function, ε represents the relative error between measurement and reconciliation as shown in Equation 57, with x^m = the measurable process variable x^r = the reconciliation rate for the process variables and σ = the standard deviation of the measurements.

$$\varepsilon = \left(\frac{x^m - x^r}{\sigma}\right) \quad (57)$$

As to define a performance criterion for function selection, Equation 58 represents aspects such as convergence and relative error reduction. The first aspect indicates when the function should be employed in real-time applications; the second refers to the ability of the function to serve in error detection (de França; de Oliveira-Júnior; de Santana-Souza, 2016).

$$RER = \frac{\sum_i (REM - RRE_i)}{\sum_i REM_i} * 100 \quad (58)$$

Where REM_i is the relative error measure and RRE_i is the reconciled relative error, as shown in Equations 59 and 60.

$$REM_i = \frac{|x_i - x_i^m|}{x_i} \quad (59)$$

$$RRE_i = \frac{|x_i - x_i^r|}{x_i} \quad (60)$$

Where x_i is the true range; x_i^m is the measured range and x_i^r is the reconciled range.

ANFIS

ANFIS (adaptive neuro-fuzzy inference system) is a kind of adaptive multilayer feed-forward network. It integrates the linguistic expression function of fuzzy inference with the self-learning characteristic of an ANN. The ANFIS system consists of two inputs and one output as shown in Figure 7, with two fuzzy if-then rules, which adjust the constants. It resembles having two networks overlapping with an initial classification that allows modeling for the same system multiple inputs of the same variable. This adjustment allows the fuzzy systems to learn from the data they are modeling (Li; Niu; Xiao, 2012; Wang *et al.*, 2021). This network is normally used over other models to improve the accuracy as sawn in Li, Niu, Liu *et al.* (2012) with an ELM application.

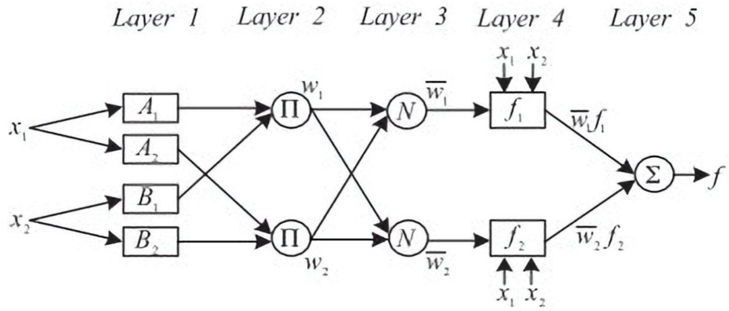

Figure 7. Architecture ANFIS.
Source: Li *et al.* (2012).

The fuzzy if-then rules are: (1) if x_1 is A_1 and x_2 is B_1 then $f_1 = p_1 x_1 + q_1 x_2 + r_1$, and (2) *if* x_1 is A_2 and x_2 is B_2 *then* $f_2 = p_2 x_1 + q_2 x_2 + r_2$, where p_1, q_1, p_2, q_2, r_1 and r_2 are constants.

RSM

Response Surface Methodology (RSM) is used to train neural networks by exploring the relationship between input variables and one or more output variables. RSM methodology is a collection of statistical and mathematical techniques applied to create empirical models. Its purpose is to optimize the output variable concerning the input variables. Regarding boilers, the most common model is one of two input variables and one output variable that fits a typical model with two input variables for the RSM methodology, as shown in Equation 61.

$$\mathrm{H}(x,y) = c_0 + c_1 x + c_2 y + c_{12} xy + c_{11} x^2 + c_{22} y^2 \qquad (61)$$

Where, c_i are constant coefficients to be found to solve the model, in this case, efficiency as an objective function, and flow rate and temperature as independent variables (Maddah *et al.*, 2019).

3. Discussion

The main objective of the exposed methodologies is to estimate the boiler's efficiency, applying models (AI or mathematical modeling) that allow minimizing fuel consumption while maximizing steam production. On the other hand, some models focus on estimate the concentration of NOx pollutants, important to validate environmental regulations. There are studies that, using methods such as thermographic analysis, examine the different boiler's energetic losses and their possible mitigation. Also, some studies analyze boiler's behavior at a macro level to define the maximum efficiency and the best combination of variables that improve their performance.

Bringing the discussion to the industrial application, it is worth evaluating the use of the different models in real-time estimation of efficiency. Mechanistic models are normally carried out in academic practice and their application at an industrial level is quite complicated, which is why empirical models emerge as an alternative, within which, historically, mathematical modeling stands out.

Taking into account that boilers are complex systems composed of several sub-systems, although most of the methodologies presented analyze the whole system, some of them specialize in those sub-systems, this is the case of NTU & LMTD that focuses on the heat transfer in the boiler's or economizer's exchanger; the thermographic analysis focuses on energy losses due to heat transfer to the environment, and is usually used to determine the integrity of the boiler insulation; and FEM, which is mechanistic modeling of the heat exchange between fluids.

Mathematical modeling is one of the most used in the industry since it does not require great computational capacity, but given the recent advances in this area, lately, it has been active work on the implementation of AI algorithms, it can relate as they constitute more than half of the exposed methodologies, where neural networks stand out, and many of the methodologies are destined to optimize or improve the precision of existing models, building hybrid models, which are no other than combinations of two or more methods. For example, Li, Niu, Liu *et al.* (2012) use ELM to obtain empirical relation, ANFIS to improve the accuracy of the model, and ABC to optimize the ELM model.

No method allows us to apply without data, that's why at the end measurements of the different operating variables of the boiler are needed, or, failing that, reliable historical data, from which the mass and energy balance, i.e., analytical methods, provide an efficiency estimation. But, measuring the variables has an associated cost, so that is the main purpose of the empirical models, for a relatively small set of variables, even a pair, define the efficiency behavior.

Once the model is adjusted, there are certain applications to improve efficiency, most common is to change the input variables to obtain, theoretically, a better efficiency, but with the existent technology, this process is a problem of optimization, giving certain values of the input variables that can be changed in practice to improve the boiler's performance.

4. Conclusions

The detailed review of the methodologies for the calculation of energy efficiency and boiler performance shows that the choice of the appropriate methodology depends on the specific case being treated. However, there is a strong preference for analytical models for a first approximation, and for systems to be described accurately, mechanistic models are more commonly used.

Analytical models are the main basis for efficiency calculations at the industrial level, since, with a starting database, exceptional results are achieved by describing the behavior of the variables of interest, such models can be based on regressions or artificial intelligence algorithms.

Artificial intelligence is a field in permanent development so that different methodologies and improvement proposals to the existing ones are emerging, where a combination of them is possible to optimize more and more the adjustment to the systems to be described, covering their shortcomings among them.

The main contribution of this paper is to generate a broad review of the existing methodologies and expose them mathematically, so that, when reviewing the article, regarding specific needs for estimate boiler's energy efficiency (data availability), it is possible to have a good approximation of which methodology works best in each case.

5. Future directions

Future work in boiler's efficiency modeling could be the application in a real-time efficiency estimation, in which the empirical methods or AI algorithms stand out. Another field of research that can be developed is the implementation of methods that automatically identify possible changes in variables to improve energy efficiency, and that closely compare these possible changes to produce suggestions whose implementation is feasible, guiding operators' decision making. Also, it is important to achieve better reliability and accuracy achieved in the models and to model the energy efficiency using fewer input variables.

Acknowledgments

This work was financed with resources from SENA and Minciencias through the Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación, Francisco José de Caldas, under contract 80740-415-2020. The Centro Industrial de Mantenimiento Integral (CIMI) in Santander participates in the transfer plan of this project.

References

Abubakar, Haruna; Bello, Ademola; Ejilah, Robinson (2020) Development of Mathematical Model of a Utility Boiler Based on Energy and Exergy Analysis. *SSRG International Journal of Thermal Engineering (SSRG-IJTE), 6*(1), 1-10. https://doi.org/10.14445/23950250/IJTE-V6I1P101

Abroshan, Hamid (2020). An integrated model to study the effects of operational parameters on the performance and pollutant emissions in a utility boiler. *Journal of Thermal Engineering*, *6*(4), 474-498. https://doi.org/10.18186/THERMAL.764157

Amell-Arrieta, Andrés; Vélez-Rueda, Liliana (2003). Análisis comparativo de eficiencias de combustión en un generador de vapor cuando se utiliza gas natural y diesel de caldera. *Revista Facultad de Ingeniería (Universidad de Antioquia)*, 28, 62-75.

American Society Mechanical Engineering (2013). *PTC 4-2013 Fired Steam Generators.* https://www.asme.org/codes-standards/find-codes-standards/ptc-4-fired-steam-generators

Apaza, Ubaldo; Delgado, Aldo; Garcilazo, Iván; Obregón, Igor (2017). Sistema de Automatización de un Caldero de 30 Bhp para el Mejoramiento de la Eficiencia Energética Utilizando Variables Termodinámicas. *Revista Científica Tecnia, 24*(1), 5-11. https://doi.org/10.21754/tecnia.v24i1.26

Bahadori, Alireza; Vuthaluru, Hari (2010). Estimation of energy conservation benefits in excess air-controlled gas-fired systems. *Fuel Processing Technology, 91*(10), 1198-1203. https://doi.org/10.1016/j.fuproc.2010.03.033

Barma, M. C.; Saidur, R.; Rahman, S. M. A.; Allouhi, A.; Akash, B. A.; Sait, S. M. (2017). A review on boilers energy use, energy savings, and emissions reductions. *Renewable and Sustainable Energy Reviews*, *79*(May), 970-983. https://doi.org/10.1016/j.rser.2017.05.187

Behbahaninia, A.; Ramezani, S.; Lotfi Hejrandoost, M. (2017). A Loss Method for Exergy Auditing of Steam Boilers. *Energy*, *140*(1). 253-260. https://doi: 10.1016/j.energy.2017.08.090

Bengio, Yoshua; LeCun, Yan (2019). Scaling Learning Algorithms toward AI. In L. Bottou; O. Chapelle; D. DeCoste; J. Weston (Eds.). *Large-Scale Kernel Machines* (pp. 1-41). MIT Press. https://doi.org/10.7551/mitpress/7496.003.0016

Bujak, Janusz (2008). Mathematical modelling of a steam boiler room to research thermal efficiency. *Energy*, *33*(12), 1779-1787. https://doi.org/10.1016/j.energy.2008.08.004

Chen, Bin; Ye, Xiao; Shen, Jun; Wang, Sha; Deng, Shenxiang; Yang, Jinbiao (2021). Investigations on the energy efficiency limits for industrial boiler operation and technical requirements- taking China's Hunan province as an example. *Energy, 220,* 119672. https://doi.org/10.1016/j.energy.2020.119672

Dedovic, Nebojsa; Igic, Sasa; Janic, Todor; Matic-Kekic, Snezana; Ponjican, Ondrej; Tomic, Milan; Savin, Lazar (2012). Efficiency of small scale manually fed boilers-mathematical models. *Energies, 5*(5), 1470-1489. https://doi.org/10.3390/en5051470

de França, Regina; de Oliveira-Júnior, Antônio; de Santana Souza, Domingos (2016). Avaliação de funções robustas para reconciliação de dados em sistemas térmicos. *Acta Scientiarum. Technology*, *38*(2), 185-191. https://doi.org/10.4025/actascitechnol.v38i2.28188

Díez, Luis; Cortés, Cristóbal; Campo, Antonio (2005). Modelling of pulverized coal boilers: Review and validation of on-line simulation techniques. *Applied Thermal Engineering*, *25*(10), 1516-1533. https://doi.org/10.1016/j.applthermaleng.2004.10.003

Ding, Yukun; Liu, Jinlang; Xiong, Jinjun; Jiang, Meng; Shi, Yiyu (2018). Optimizing boiler control in real-time with machine learning for sustainability. In A. Cuzzocrea (Chair), *Proceedings from the International Conference on Information and Knowledge Management* (pp. 2147-2154). Association for Computing Machinery. https://doi.org/10.1145/3269206.3272024

Dorotić, Hrvoje; Pukšec, Tomislav; Duić, Neven (2020). Analysis of displacing natural gas boiler units in district heating systems by using multi-objective optimization and different taxing approaches. *Energy Conversion and Management*, *205*, 112411. https://doi.org/10.1016/j.enconman.2019.112411

Farhat, Alaa; Zoughaib, Assaad; El Khoury, Khalil (2015). A new methodology combining total site analysis with exergy analysis. *Computers and Chemical Engineering, 82,* 216-227. https://doi.org/10.1016/j.compchemeng.2015.07.010

Forkman, Johannes ; Josse, Julie; Piepho, Hans-Peter (2019). Hypothesis Tests for Principal Component Analysis When Variables are Standardized. *Journal of Agricultural, Biological, and Environmental Statistics*, *24*(2), 289-308. https://doi.org/10.1007/s13253-019-00355-5

García-Sánchez, Gabriel; Chacón-Velasco, Jorge; Fuentes-Díaz, David; Jaramillo-Ibarra, Julián; Martínez-Morales, Jairo (2020). Modelado CFD de la combustión en calderas de biomasa – Revisión del estado del arte. *Respuestas*, *25*(3), 262-273. https://doi.org/10.22463/0122820X.2462

Granados, Roberto (2016). *Modelos de regresión lineal múltiple. Documentos de Trabajo en Economía Aplicada*. Universidad de Granada.

Gutiérrez Ortiz, Francisco (2011). Modeling of fire-tube boilers. *Applied Thermal Engineering*, *31*(16), 3463-3478. https://doi.org/10.1016/j.applthermaleng.2011.07.001

Hasini, Hasril; Yusoff, Mohd; Shuaib, Norshah; Boosroh, Mohd; Haniff, Muhammad (2009). Analysis of flow and temperature distribution in a full-scale utility boiler using CFD. *ICEE 2009. Proceedings of the 3rd International Conference on Energy and Environment: Advancement Towards Global Sustainability*, *2*(December), 208-214. https://doi.org/10.1109/ICEENVIRON.2009.5398646

Irwin, G.; Brown, G.; Hogg, B.; Swidenbank, E. (1995). Neural network modelling of a 200MW boiler system. *IEE Proceedings: Control Theory and Applications*, *142*(6), 529-536. https://doi.org/10.1049/ip-cta:19952293

Ivanitckii, M. S.; Sultanov, M. M.; Kuryanova, E. V. (2021) Development of a mathematical model for evaluating the technical condition of boiler plants in terms of reliability and efficiency. In *IEE 2021 3rd International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)* (pp. 1-4). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/REEPE51337.2021.9388050

Jiménez-Borges, Reiner; Madrigal, José; Lapido, Margarita; Vidal, David (2016). Método para la evaluación de la eficiencia e impacto ambiental de un generador de vapor. *Ingeniería Energética, 37*(2),135-143.

Kaewboonsong, W.; Kuprianov, V. I. ; Chovichien, N. (2006). Minimizing fuel and environmental costs for a variable-load power plant (co-)firing fuel oil and natural gas. Part 1. Modeling of gaseous emissions from boiler units. *Fuel Processing Technology*, *87*(12), 1085-1094. https://doi.org/10.1016/j.fuproc.2006.08.003

Kær, Søren (2004). Numerical modelling of a straw-fired grate boiler. *Fuel, 83*(9), 1183-1190. https://doi.org/10.1016/j.fuel.2003.12.003

Kerr, D.; Blair, A. D. (2011). Thermal Power Plant Performance Analysis. *Physiotherapy (United Kingdom), 97*. https://doi.org/10.1016/j.physio.2011.04.002

Kim, Jongho; Lee, Soonho; Tahmasebi, Arash; Jeon, Chun-Hwan; Yu, Jianglong (2021) A Review of the Numerical Modeling of Pulverized Coal Combustion for High-Efficiency, Low-Emissions (HELE) Power Generation. *Energy & Fuels, 35*(9), 7434-7466. https://doi.org/10.1021/acs.energyfuels.1c00343

Kljajić, Miroslav; Gvozdenac, Dušan; Vukmirović, Srdjan (2012). Use of Neural Networks for modeling and predicting boiler's operating performance. *Energy*, *45*(1), 304-311. https://doi.org/10.1016/j.energy.2012.02.067

Lang, Fred (2009). *Errors in boiler efficiency standards.* In *Proceedings of ASME Power Conference* (pp. 487-501). American Society of Mechanical Engineers. https://doi.org/10.1115/POWER2009-81221

Lawrence, S.; Giles, C. L.; Tsoi, A. C. (1996). *What size neural network gives optimal generalization? Convergence properties of backpropagation* (Technical Report UMIACS-TR-96-22 and CS-TR-3617). Institute for Advanced Computer Studies, University of Maryland.

Li, Guoqiang; Niu, Peifeng; Liu, Chao; Zhang, Weiping (2012). Enhanced combination modeling method for combustion efficiency in coal-fired boilers. *Applied Soft Computing Journal*, *12*(10), 3132-3140. https://doi.org/10.1016/j.asoc.2012.06.016

Li, Guoqiang; Niu, Peifeng; Xiao, Xingjun (2012). Development and investigation of efficient artificial bee colony algorithm for numerical function optimization. *Applied Soft Computing Journal*, *12*(1), 320-332. https://doi.org/10.1016/j.asoc.2011.08.040

Lozano, M. A.; Valero, A. (1993). Theory of exergetic cost. *Energy*, *18*(9), 939-959.

Maddah, Heydar; Sadeghzadeh, Milad; Ahmadi, Mohammad; Kumar, Ravinder; Shamshirband, Shahaboddin (2019). Modeling and efficiency optimization of steam boilers by employing neural networks and response-surface method (RSM). *Mathematics*, *7*(7). https://doi.org/10.3390/math7070629

Math, Praveen; Chandrakar, Tikendra; Kumar, Santosh; Bhamniya, Jaiprakash; Rinawa, Moti-Lal; Ankit (2021). Experimental and CFD simulation of performance analysis of steam generators of boilers. *Materials Today: Proceedings*, *46*(17), 7049-8894. https://doi.org/10.1016/j.matpr.2021.07.084

Milićević, Aleksandar; Belošević, Srdjan; Crnomarković, Nenad; Tomanović, Ivan; Tucaković, Dragan (2020). Mathematical modelling and optimisation of lignite and wheat straw co-combustion in 350 MW boiler furnace. *Applied Energy*, *260*, 114206. https://doi.org/10.1016/j.apenergy.2019.114206

Modliński, Norbert; Szczepanek, Krzysztof; Nabagło, Daniel; Madejski, Paweł; Modliński, Zbigniew (2019). Mathematical procedure for predicting tube metal temperature in the second stage reheater of the operating flexibly steam boiler. *Applied Thermal Engineering*, *146*, 854-865. https://doi.org/10.1016/j.applthermaleng.2018.10.063

Qu, Ming; Abdelaziz, Omar; Yin, Hongxi (2014). New configurations of a heat recovery absorption heat pump integrated with a natural gas boiler for boiler efficiency improvement. *Energy Conversion and Management, 87*, 175-184. https://doi.org/10.1016/j.enconman.2014.06.083

Qin, Na; Li, Rui (2020) Online simplified model and experimental comparison of CFB boiler thermal efficiency. *Applied Thermal Engineering, 171*, 115021. https://doi.org/10.1016/j.applthermaleng.2020.115021

Rehan, Ahmed; Habib, Mohamed; Elshafei, Moustafa; Alzaharnah, Iyad (2018). Modeling Time Variations of Boiler Efficiency. *Journal of Energy Resources Technology, Transactions of the ASME*, *140*(5), 1-58. https://doi.org/10.1115/1.4038236

Retirado-Mediaceja, Yoalbys; Camaraza-Medina, Yanán; Sánchez-Escalona, Andrés; Laurencio-Alfonso, Héctor; Salazar-Corrales, Marcelo; Zalazar-Oliva, Carlos (2020). Thermo-exergetic assessment of the steam boilers used in a cuban thermoelectric facility. *International Journal of Design and Nature and Ecodynamics*, *15*(3), 291-298. https://doi.org/10.18280/ijdne.150302

Rusinowski, Henryk; Stanek, Wojciech (2007). Neural modelling of steam boilers. *Energy Conversion and Management*, *48*(11), 2802-2809. https://doi.org/10.1016/j.enconman.2007.06.040

Rusinowski, Henryk; Stanek, Wojciech (2010). Hybrid model of steam boiler. *Energy*, *35*(2), 1107-1113. https://doi.org/10.1016/j.energy.2009.06.004

Saha, P. K.; Shoib, Mohammed; Kamruzzaman, J. (1998). Development of a neural network based integrated control system of 120 ton/h capacity boiler. *Computers and Electrical Engineering*, *24*(6), 423-440. https://doi.org/10.1016/S0045-7906(98)00020-2

Sankar, G.; Santhosh-Kumar, D.; Balasubramanian, K. R. (2019) Computational modeling of pulverized coal fired boilers – A review on the current position. *Fuel, 236,* 643-665. https://doi.org/10.1016/j.fuel.2018.08.154

Savargave, Sangram; Lengare, Madhukar (2017). Self-Adaptive Firefly Algorithm with Neural Network for Design Modelling and Optimization of Boiler Plants. In *International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 289-293). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/I-SMAC.2017.8058357

Savargave, Sangram; Lengare, Madhukar (2018). Modeling and Optimizing Boiler Design using Neural Network and Firefly Algorithm. *Journal of Intelligent Systems*, *27*(3), 393-412. https://doi.org/10.1515/jisys-2016-0113

Serway, Raymond; Jewett, John (2008). *Física para Ciencias e Ingeniería* (Vol. 1, 7ª ed.). Cengage Learning.

Szega, Marcin (2020). Methodology of advanced data validation and reconciliation application in industrial thermal processes. *Energy*, *198*, 117326. https://doi.org/10.1016/j.energy.2020.117326

Szega, Marcin; Nowak, Grzegorz (2015). An optimization of redundant measurements location for thermal capacity of power unit steam boiler calculations using data reconciliation method. *Energy*, *92*, 135-141. https://doi.org/10.1016/j.energy.2015.03.125

Taler, Jan; Dzierwa, Piotr; Taler, Dawid; Harchut, Piotr (2015). Optimization of the boiler start-up taking into account thermal stresses. *Energy*, *92*, 160-170. https://doi.org/10.1016/j.energy.2015.03.095

Tang, Zhenhao; Li, Yanyan; Kusiak, Andrew (2020). A Deep Learning Model for Measuring Oxygen Content of Boiler Flue Gas. *IEEE Access, 8,* 12268-12278. https://doi.org/10.1109/ACCESS.2020.2965199

Tarasevich, Maksimilian; Tepljakov, Aleksei; Petlenkov, Eduard; Vansovits, Vitali (2020). Modeling and Identification of an Industrial Hot Water Boiler. In *IEEE 2020 43rd International Conference on TSP* (pp. 285-290). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/TSP49548.2020.9163503

Tognoli, Marco; Najafi, Behzad; Rinaldi, Fabio (2018). Dynamic modelling and optimal sizing of industrial fire-tube boilers for various demand profiles. *Applied Thermal Engineering*, *132*, 341-351. https://doi.org/10.1016/j.applthermaleng.2017.12.082

Trojan, Marcin (2019). Modeling of a steam boiler operation using the boiler nonlinear mathematical model. *Energy*, *175*, 1194-1208. https://doi.org/10.1016/j.energy.2019.03.160

Unidad de Planeación Minero Energética; Institute for Resource Efficiency and Energy Strategies; TEP Energy; Corpoema (2019). *Primer balance de energía útil para Colombia y cuantificación de las pérdidas energéticas relacionadas y la brecha de eficiencia energética. Resumen ejecutivo BEU sector industrial.* Unidad de Planeación Minero Energética.

Valencia-Ochoa, Guillermo; Rojas, Jhan; Campos-Avella, Juan. (2019). Energy optimization of industrial steam boiler using energy performance indicator. *International Journal of Energy Economics and Policy*, *9*(6), 109-117. https://doi.org/10.32479/ijeep.8188

Wang, Y. F.; Wang, M. X.; Liu, Y.; Yin, L.; Zhou, X. R.; Xu, J. F.; Zhang, X. Y. (2021). Fuzzy modeling of boiler efficiency in power plants. *Information Sciences, 542,* 391-405. https://doi.org/10.1016/j.ins.2020.06.064

Whitley, Darrell (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65-85. https://doi.org/10.1007/BF00175354

Zhang, Ruiqing; Yang, Hairui; Hu, Nan; Lu, Junfu; Wu, Yuxin (2013). Experimental investigation and model validation of the heat flux profile in a 300MW CFB boiler. *Powder Technology*, *246*, 31-40. https://doi.org/10.1016/j.powtec.2013.04.038

Zhang, Yi; Ding, Yanjun; Wu, Zhansong; Kong, Liang; Chou, Tao (2007). Modeling and coordinative optimization of NO x emission and efficiency of utility boilers with neural network. *Korean Journal of Chemical Engineering*, *24*(6), 1118-1123. https://doi.org/10.1007/s11814-007-0131-6

Zhao, Jing; Duan, Yaoqi; Liu, Xiaojuan (2019). Study on the policy of replacing coal-fired boilers with gas-fired boilers for central heating based on the 3E system and the TOPSIS method: A case in Tianjin, China. *Energy*, *189*, 116206.
https://doi.org/10.1016/j.energy.2019.116206

Zhitarenko, Volodymyr; Bejan, Volodymyr; Ostapenko, Oleksij (2020). Adaptation of Mathematical Model of Heat and Energy Characteristics of Medium Pressure Boilers to Real Operating Conditions. *Technology Audit and Production Reserves, 4*(1), 23-30.
https://doi.org/10.15587/2706-5448.2020.210540

Zhou, Jinjun; Deng, Song; Turner, Dan; Claridge, David; Haberl, Jeff (2002). *Improving boiler efficiency modeling based on ambient air temperature*. Proceedings of the Thirteenth Symposium on Improving Building Systems in Hot and Humid Climates, Houston, TX.