Comparing a linear transfer function-noise model and a neural network to model boiler bank fouling in a kraft recovery boiler

JERRY NG, GUSTAVO M. DE ALMEIDA, ESA K. VAKKILAINEN, YURI A. LARYSHYN, AND NIKOLAI A. DEMARTINI

ABSTRACT: Boiler bank fouling reduces heat transfer efficiency in kraft recovery boilers. Here, we model the relationships between boiler parameters and boiler bank pressure drop, an indicator of fouling, based on recovery boiler operating data. We compared two models: an autoregressive integrated exogenous (ARIX) model and a feed-forward neural network. The ARIX model better simulates boiler bank pressure drop compared to the neural network (R^2 of 0.64 vs. 0.58). Based on the ARIX model, we identified six boiler parameters that significantly influence boiler bank fouling and their relative contributions. Finally, we demonstrate how the models can simulate boiler bank pressure drop given artificial perturbations in boiler parameters.

Application: This study shows that a significant amount of variability in boiler bank fouling can be modeled with linear models, which can lead to new control strategies. Additionally, those looking to apply neural networks to mill data can use the procedures in this study to properly train and benchmark neural network models.

In the kraft process, wood chips are digested in white liquor, an alkaline solution that contains sodium hydroxide (NaOH) and sodium sulfide (Na_2S) as active chemicals. The digestion of wood dissolves roughly half of the wood and frees fibers that compose pulp. Dissolved organics and inorganic pulping chemicals from the white liquor are separated from the pulping, forming an aqueous solution called black liquor. Water is evaporated from the black liquor, reducing its water content from around 85% to as low as 15% [1]. After that, the black liquor is burned in a kraft recovery boiler to recover energy and the pulping chemicals [2].

In the lower furnace of the boiler, heat is transferred to the water-cooled walls of the boiler, mainly via radiation. As combustion gases flow up through the upper furnace, the hot gases pass through multiple convective heat transfer sections composed of sets of tubes in hanging platens. The hot gases enter the superheater section of the tubes at about 1000°C before entering the boiler bank section at 550°C–700°C. Finally, the gases enter the economizer section at roughly 200°C [1]. Deposits of inorganic salts on heat transfer surfaces, called fouling, decrease heat transfer in the convective zones of the boiler, reducing boiler efficiency. Fouling is controlled by manipulating boiler operating parameters and by deposit removal using sootblowers.

In kraft recovery boilers, inorganics reach heat transfer surfaces in three forms:

1. *Carryover large (0.01–3 mm)*: molten smelt particles and partially burned black liquor solids entrained in the combustion gases that flow to the upper furnace.
2. *Intermediate sized particles (ISP)*: smaller (1–100 µm) molten or partially molten smelt particles from material ejecting from burning black liquor droplets in the lower furnace and carried up with the combustion gases to the upper furnace.
3. *Fume*: the smallest particles (0.1–1 µm) resulting from the condensation of sodium and potassium compound vapors released from burning black liquor droplets.

Carryover particles dominate deposit formation at the superheaters and at the front of the boiler bank, while fume is the primary source of deposits in the economizers. Intermediate sized particles represent a smaller fraction of the total inorganics in the upper furnace but typically form deposits in the superheaters and boiler bank [3,4].

In this work, we predict fouling in the boiler bank based on boiler operating parameters and upstream measurements. Fouling at the front of the boiler bank, closer to the furnace, is mainly due to carryover and ISP. Downstream, towards the back of the boiler bank, the contribution of fume to deposit formation eclipses those of carryover and ISP. At the back of the boiler bank, temperatures are lower (500°C), resulting in vapor condensa-

tion, and most carryover and ISP have been screened out by the upstream tubes.

Boiler bank deposits formed near the superheater create a fouling "front" that travels towards the back of the boiler bank. Deposits reduce the ability of incoming flue gas to be cooled, so, over time, high temperature deposits form further downstream in the boiler bank [5]. Increased deposition results in increased steam costs for sootblowing.

A draft fan draws the combustion gases from the lower furnace through the convective passes and out the stack. As deposits accumulate on heat transfer surfaces, the channels through which the combustion gases travel become constricted, creating a pressure drop across the boiler bank. As a result, the load on the draft fan increases and eventually reaches its maximum. This prevents the boiler from operating at full load and may require an unplanned outage to perform water washing to remove excessive deposit buildup. Ideally, if the relationship between boiler parameters and fouling were well understood, fouling could be mitigated, delaying water washing until planned outages.

Boiler bank fouling is known to be affected by liquor firing rate, firing temperature, firing pressure, liquor properties, air distribution, and lower furnace conditions [6]. However, some variables such as boiler properties and droplet size distribution are not measured in real time. Reconciling these unmeasured disturbances with other boiler parameters is difficult; as a result, periods of high fouling can be difficult to predict.

Additionally, boiler shutdowns can affect the process data. When the boiler heat transfer surfaces are cleaned, the fouling starts anew. Typically, the beginning of this new fouling is faster, lasting from a few days to two weeks. Deposits formed during this period are fairly cold and are thus stronger. As the deposits continue to regrow, the outer layers are hotter and thus easier to remove [1].

The impact of such periods on the resulting models or analyses would also depend on the duration of the boiler shutdowns relative to the rest of the dataset. If the duration of the shutdown is relatively small, the effect on model parameters and results may be negligible. In contrast, if the shutdown periods, or other periods of erroneous data, constitute a significant amount of the total dataset, the impact may be large. Modeling approaches that are more interpretable or where feature analysis is more rigorous should be more robust to such errors and identify effects more easily.

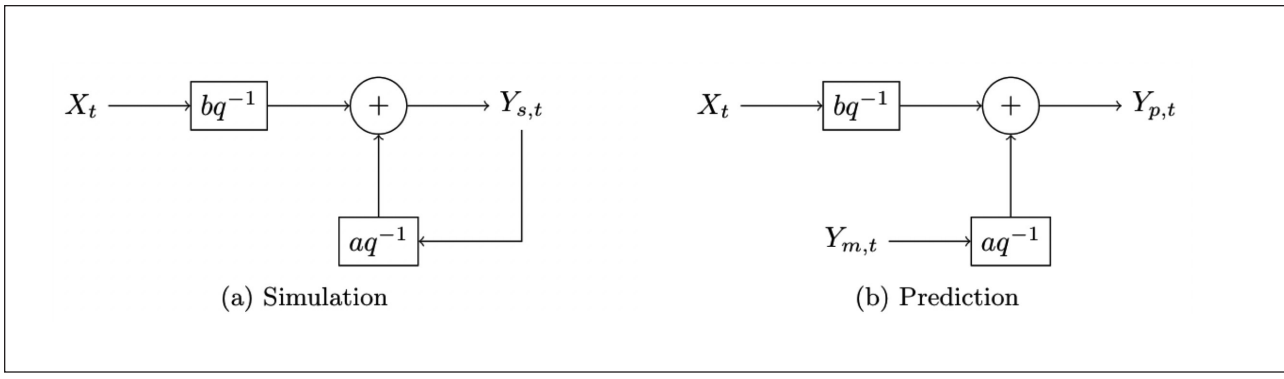
Some studies have used computational fluid dynamics to model deposit formation and have compared their predictions to boiler measurements with reasonable success [7]. This approach is based on first-principles and is a powerful tool when considering changes in air distribution or air box design, for example. However, such models are typically impractical to implement in daily operations, as they are computationally expensive, require extensive expertise to modify, and may not account for process noise.

MACHINE LEARNING MODELS AND PREVIOUS APPLICATIONS TO KRAFT RECOVERY BOILERS

In contrast to first-principles models, machine-learning models approximate a phenomenon based on measured data. In the past, there was a heavy preference for such models to be parsimonious and robust [8]. As a result, linear models that were dynamic and stochastic were favored [9,10]. Additionally, models were mostly judged on their usefulness rather than whether they captured the "truth" of a system [11]. These principles were partly a result of constrained software and computing power. Recently, however, increasingly complex machine-learning models have become more accessible, particularly neural networks. Now, large, sophisticated nonlinear models are commonly proposed for industrial processes [12-16], including kraft recovery boilers [17-20].

Example applications of machine learning to recovery boiler fouling include Versteeg and Tran [21], who used principal component analysis (PCA) to identify kraft recovery boiler parameters related to high and low fouling regimes, as well as Belisario et al. [22], who used a neural network to correlate kraft recovery boiler parameters with measured fouling. In both cases, the models were reported to explain significant amounts of variability in the observed fouling. The PCA analyses by Versteeg and Tran explained 64%–69% of observed variability across three boilers; the neural network trained by Belisario et al. achieved an R^2 of 0.85 on unseen test data when predicting fouling in a single boiler. These results are important, as they encourage boiler fouling management to move from heuristics to more systematic approaches, and they incentivize plants to improve and maintain process measurements.

Unfortunately, neither work accounted for the dynamics of process data [23,24]. Belisario et al. shuffled hourly data before separating it into training and testing sets. As a result, the training and testing data would be highly correlated. Versteeg and Tran used daily averaged data, so dynamic relationships would not be accounted for. Additionally, while cross validation was used to avoid overfitting, it is unclear if the data was shuffled. If not, the training and validation sets would be highly correlated, inflating model performance. Similarly, Romeo and Gareta [25-27] paired a neural network model and a fuzzy logic control system to minimize fouling in a biomass boiler. However, the neural network was trained and validated with randomly shuffled data. Again, this misrepresents the model performance, because the training and validation data are likely highly correlated. Finally, all of these studies rely on black box models that are difficult to interpret or implement, and the black box approaches were not compared to simpler, classic dynamic models. In the literature, considering autocorrelation and benchmarking against simpler models is commonly ignored, and, perhaps as a result, there is increasing evidence that the usefulness of complex



1. Simulation vs. prediction.

machine-learning models relative to simpler models for dynamical systems may be exaggerated [28-32].

Simulation vs. prediction

Figure 1 illustrates the difference between simulation and prediction of a model characterized by the parameters a and b. The shift operator q operates such that $q^{-1}Y_t = Y_{t-1}$. In simulation (Fig. 1a), past model outputs are used to generate new model outputs. In contrast, in prediction (Fig. 1b), past measurements of the output are used to generate new model outputs. Importantly, Figs. 1a and 1b describe the same model structure, but the manner in which the outputs are generated is different.

Typically, models are trained to minimize prediction error $Y_t = Y_{m,t}$. However, models trained to minimize prediction error can learn to rely entirely on past output measurements and ignore exogenous inputs: there is a "path" to the output Y_t that ignores X_t. This is likely to occur if the output is autocorrelated, as is common with process data. In contrast, models trained to minimize simulation error $Y_t = Y_{s,t}$ are forced to learn a path from $X_t = Y_t$.

Noise models

Noise models are models that only comprise endogenous inputs. The persistence model and mixed autoregressive moving average (ARMA) model are examples of noise models [10]. Assuming a white noise process $e_t \sim N(0, \sigma^2)$, the persistence model and ARMA model can be expressed as,

$$Y_t = \frac{1}{1-q^{-1}} e_t$$

and

$$Y_t = \frac{1+\theta_1 q^{-1}+\theta_2 q^{-2}+\cdots+\theta_{n_\theta} q^{-n_\theta}}{1-\phi_1 q^{-1}-\phi_2 q^{-2}-\cdots-\phi_{n_\phi} q^{-n_\phi}} e_t = \frac{\theta(q)}{\phi(q)} e_t,$$

respectively, where n_θ and n_ϕ are the degrees of polynomials $\theta(q)$ and, $\phi(q)$, respectively. These models are called noise models because their only input is a noise process e_t. While these models may accurately predict Y_t, they are blind to the inputs driving Y_t. Instead, the aggregate effect of all inputs is approximated as white noise e_t.

Black box models that are trained to minimize prediction error based only on endogenous inputs are noise models. Additionally, black box models trained to minimize prediction error based on endogenous and exogenous inputs may still approximate noise models: the model may learn to ignore the exogenous inputs. Though a black box model is difficult to interpret, feature importance techniques can reveal which input a black box model relies on most heavily. If the reliance on endogenous inputs dominates, the black box model likely approximates a noise model. Training the model to minimize simulation error instead prevents the learning of a pure noise model, because the model is forced to learn a path between exogenous inputs and the output.

Training black box models to minimize simulation error

The conventional method of training black box models, such as neural networks and random forests, is to minimize prediction error, not simulation error. If the model inputs exclude endogenous inputs, simulation error and prediction error are equal. However, this is often not the case: lagged output measurements are usually given to the model to boost prediction accuracy.

To minimize simulation error, the black box model could be trained with only exogenous inputs. However, this carries a similar risk to implementing a pure feed-forward control scheme: one is unlikely to identify all relevant inputs. An alternative method is to train the model to recursively predict the entire sequence of training data, as depicted in Fig. 1a. The calculation of the loss function and subsequent parameter adjustment occur after the model has simulated the entire output sequence. **Figure 2** outlines this training procedure. Legaard et al. [33] detail how to train neural networks in this manner and describe possible augmentations.

Propagating model outputs forward prevents the model from learning a pure noise model, as a pure noise model cannot minimize simulation error. Models like recurrent neural networks use a similar mechanism but only for the length of the sequence that they are trained on. In contrast,

Gather training data into an $n_{\text{timesteps}} \times n_{\text{features}}$ array;
for epoch $= 1 : n_{\text{epochs}}$ **do**
  Create a copy array of the training data;
  for timestep $= 2 : n_{\text{timesteps}}$ **do**
    In the copy array, predict and replace target at timestep j based on entries at timestep $j-1$
  end
  Calculate the loss between the target in the copy array and the target in the training data;
  Calculate gradients of the loss function;
  Adjust the neural network weights based on calculated gradients;
end

2. Algorithm to train a neural network to minimize simulation error.

with the training approach described in Fig. 2, there is no constraint on the length of the sequence. Importantly, while a transfer function will be learned, the model may still learn to rely heavily on a noise model ($a \gg b$ in Fig. 1a). Training to minimize simulation error only guarantees that a transfer function is learned at all.

Combined transfer function-noise models

In classic systems identification [11] and time series analysis [10], combined transfer function-noise models are represented by:

$$A(q)Y_t = \frac{B(q)}{F(q)}X_t + \frac{C(q)}{D(q)}e_t$$

where A, B, C, D, and F are polynomials in q and X_t is an exogenous input. Typically, only a subset of these polynomials is used. For example, an autoregressive exogenous (ARX) model comprises polynomials A and B:

$$Y_t = \frac{B(q)}{A(q)}X_t + \frac{1}{A(q)}e_t$$

While this family of models is currently seldom considered for kraft recovery data or other process data, they have been applied to simulated recovery boiler data [34]. The black box models commonly used today are combined transfer function-noise models where combinations of endogenous and exogenous inputs are used to predict an output. However, unlike black box models, classic combined transfer function-noise models distinguish the noise model and the transfer function.

Objectives for this study

Our objectives are to determine whether dynamic, data-driven models can simulate boiler bank fouling and to compare the accuracies of a linear combined transfer function-noise model and a neural network for simulating boiler bank fouling.

To our knowledge, training neural networks to simulate, rather than predict, has not been demonstrated with kraft process data or other process data. Instead, in most applications of neural networks to process data, the neural networks are trained to minimize prediction error and are not benchmarked against classic system identification models. As a result, the benefit of more complex models in the literature may be exaggerated, as the autocorrelation the complex models likely rely on could be accounted for with simpler models. In contrast, here, we train a neural network in a manner that prevents overfitting to autocorrelated noise and compare it to a simpler, linear model.

METHODOLOGY

Dataset description and preprocessing

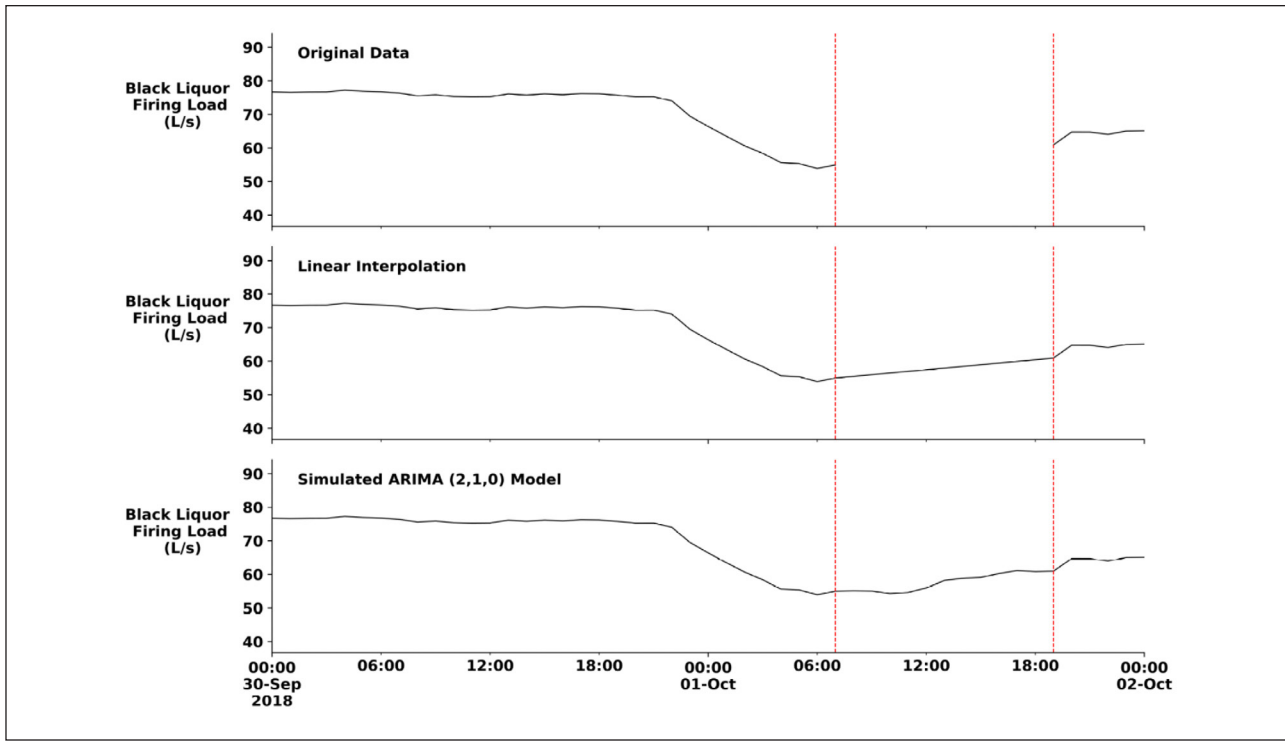
This work uses a dataset that comprises hourly measurements of 18 mill parameters from a South American mill over 496 days. First, we normalize the original flue gas pressure drop across the boiler bank ΔP_t by the steam production rate Q_t:

$$\Delta P_t^* = \Delta P_t \left(\frac{Q_t}{\overline{Q}_t}\right)^2$$

where ΔP^*_t is the normalized pressure drop and $\overline{Q}_t$ is the average steam production rate. This normalization is used in other analyses of recovery boiler fouling [35-36] and is done to account for changes in flue gas velocity. Direct flue gas velocity measurements are unreliable, so steam production rate is used to approximate flue gas velocity.

There was a small amount of missing data in the dataset (< 0.5%). Likely, most of these missing periods are due to boiler shutdowns (scheduled or unscheduled), since periods of missing data usually occur simultaneously across the boiler parameters. The ΔP_t at the end of most missing periods is lower than the beginning, suggesting that cleaning has occurred.

Two exceptions are the measurements of secondary air temperature at the rear wall and black liquor firing temperature, which contained periods of missing data not shared by the other measurements. As a result, these were likely due to instrumentation issues rather than a boiler



3. Example comparison of linear and simulated autoregressive integrated moving average (ARIMA) interpolation.

shutdown. Nevertheless, the amount of missing data is relatively small, so we do not suspect that these periods and their imputations would significantly affect models fit to the data.

Missing data was imputed with simulated autoregressive integrated moving average (ARIMA) models. Simulated ARIMA models were used to preserve the characteristic noise of each parameter. The ARIMA models were fit to the data of each mill parameter that contained missing values using the Box-Jenkins methodology [10]. Periods of missing data were imputed by simulating the ARIMA model (using random number generation) for the length of the period. While the ARIMA model parameters were determined using the entire dataset, during imputation, only historical data affects the trend of the imputation. Compared to linear interpolation, simulated ARIMA processes produce more realistic imputations (**Fig. 3**). After imputation, the unshuffled data was split into training, validation, and testing sets that comprise 60%, 15%, and 25% of the dataset, respectively.

Finally, we estimated the lag between the measurement of white liquor sulfidity and ΔP^*_t. The other mill parameters in the dataset are measured at the recovery boiler, so no lags are expected. In contrast, white liquor sulfidity is measured upstream at the digester. The lag was estimated by fitting the ARX model:

$$Y_t = \frac{B(q)}{A(q)}X_t + \frac{1}{A(q)}e_t,$$
$$A(q) = 1 - a_1 q^{-1} - a_2 q^{-2} - \cdots - a_{n_a} q^{-n_a}$$
$$B(q) = b_1 q^{-n_k} + b_2 q^{-2} + \cdots + b_{n_b} q^{-n_k - n_b}$$

where X_t is white liquor sulfidity and Y_t is ΔP^*_t. The ARX model was fit on training data, and the lag is the value of n_k, which, along with degrees n_a and n_b, minimizes validation error. For this dataset, the first change in ΔP^*_t due to a change in white liquor sulfidity is estimated to occur 33 h later, which agrees with process knowledge.

MODEL DESCRIPTIONS AND TRAINING

ARIX model

The autoregressive integrated exogenous (ARIX) model is a combined transfer function-noise model with the form:

$$Y_t = \frac{B(q)}{A(q)}X_t + \frac{1}{(1-q^{-1})A(q)}e_t$$

An ARIX model is similar to an ARX model except for the addition of an integrated noise term. We include an integrated noise term because ΔP^*_t is assumed to be a nonstationary process; if uncontrolled ($X_t = 0$) boiler bank fouling should increase over time. The ARIX model was trained to minimize simulation mean squared error (MSE):

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N is the number of samples and $\hat{y}$ is the model estimate. The optimal model hyperparameters (n_a, n_b, and n_k) were those that minimized validation MSE.

Once the ARIX model was trained, 95% confidence intervals for the model parameters (a and b) were constructed.

Additionally, the model identification process was repeated with inputs scaled to zero mean and unit variance so relative contributions could be measured. The ARIX model trained using only the statistically significant inputs performed similarly to the original ARIX model on the test data.

Feedforward neural network

To test whether considering nonlinear relationships improves the approximation of boiler bank fouling, we trained a feedforward neural network to simulate ΔP^*_t using the significant inputs identified during the training of the ARIX model. Neural networks approximate nonlinear functions through successive linear combinations and nonlinear transformations [37]:

$$\begin{aligned}\mathbf{h_1} &= \phi\left(\mathbf{w_1^T x} + \mathbf{b_1}\right),\\ \mathbf{h_2} &= \phi\left(\mathbf{w_2^T h_1} + \mathbf{b_2}\right),\\ &\cdots\\ \hat{\mathbf{y}} &= \mathbf{w_{H+1}^T h_H} + \mathbf{b_{H+1}},\end{aligned}$$

where $\mathbf{x}$ is a vector of inputs, $\mathbf{y}$ is a vector of outputs, $\mathbf{w_i}$ is a vector of weights, $\mathbf{b_i}$ is a vector of bias terms, and ϕ is a nonlinear function (e.g., tanh). The neural network used in this work comprises three fully connected layers, each with 10 hidden units that use rectified linear unit activation functions. The model was trained using backpropogation to minimize simulation MSE on the validation set for 200 epochs.

RESULTS AND DISCUSSION

Simulation accuracy

The ARIX model better simulates ΔP^*_t than the neural network model, achieving a simulation R^2 of 0.641 vs. 0.579 on the test data (**Fig. 4** and **Fig. 5**) Whereas the ARIX model tends to underpredict ΔP^*_t, the neural network model tends to overestimate ΔP^*_t by a greater degree. Despite the deviations, both models generally simulate the noise and trend of the actual data.

There are some larger deviations between model outputs and the actual ΔP^*_t that tend to persist, notably after August 20 for the ARIX model. These larger drops are likely due to brief disturbances not accounted for by the model. Since the model is run in simulation, it does not "see" the actual data, so the error between its output and the true target persists.

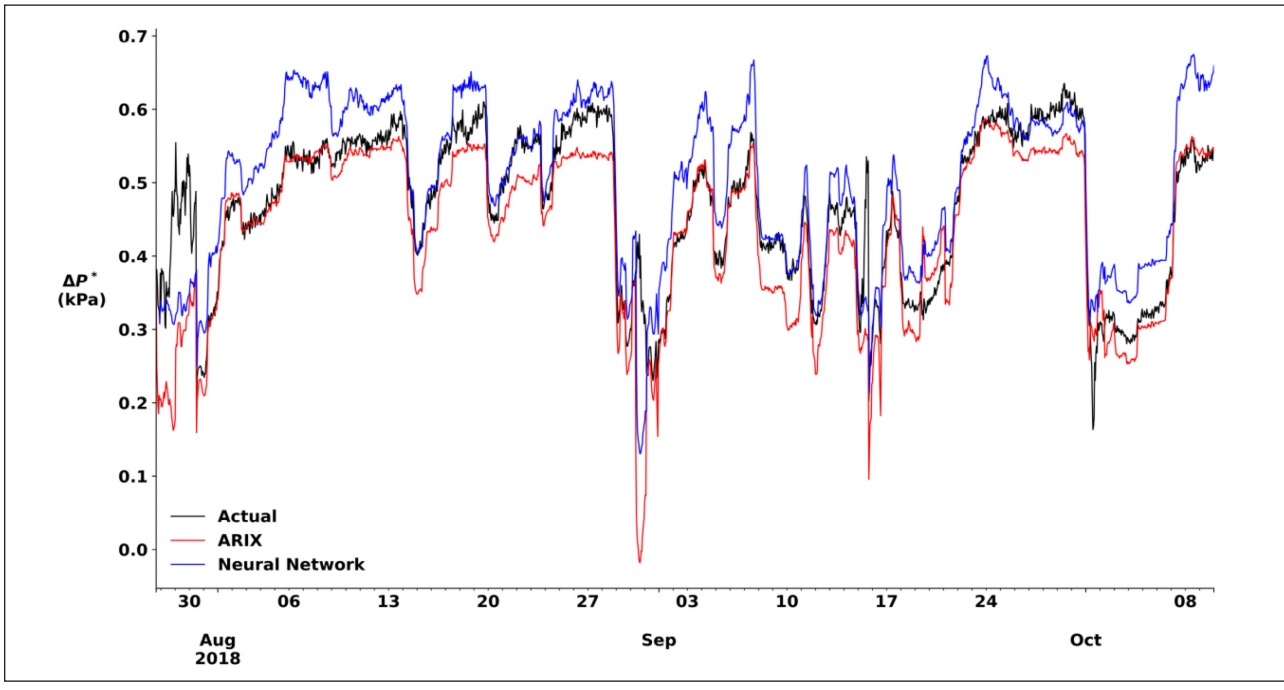
In contrast, a model that uses past output measurements would avoid this issue. However, past output measurements are not always available. For multistep predictions, such as model predictive control, model outputs must be reused. Models that are trained to minimize prediction error will rely heavily on past output measurements and will struggle when the output measurements are replaced by past model outputs.

ARIX model feature importance

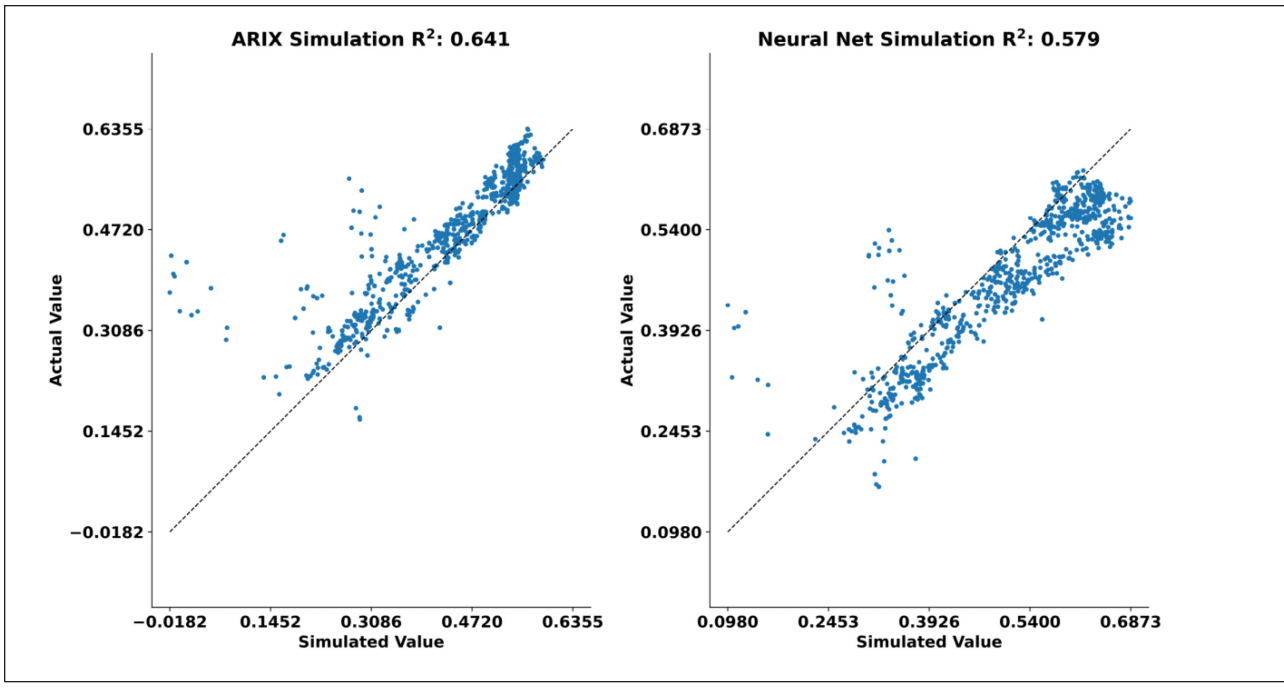
The identified ARIX model has the form:

$$\Delta P^*_t = \Delta P^*_{t-1} + \sum_{j=1}^{m} b_j q^{-n_{k,j}}\left(X_{j,t} - X_{j,t-1}\right) + e_t$$

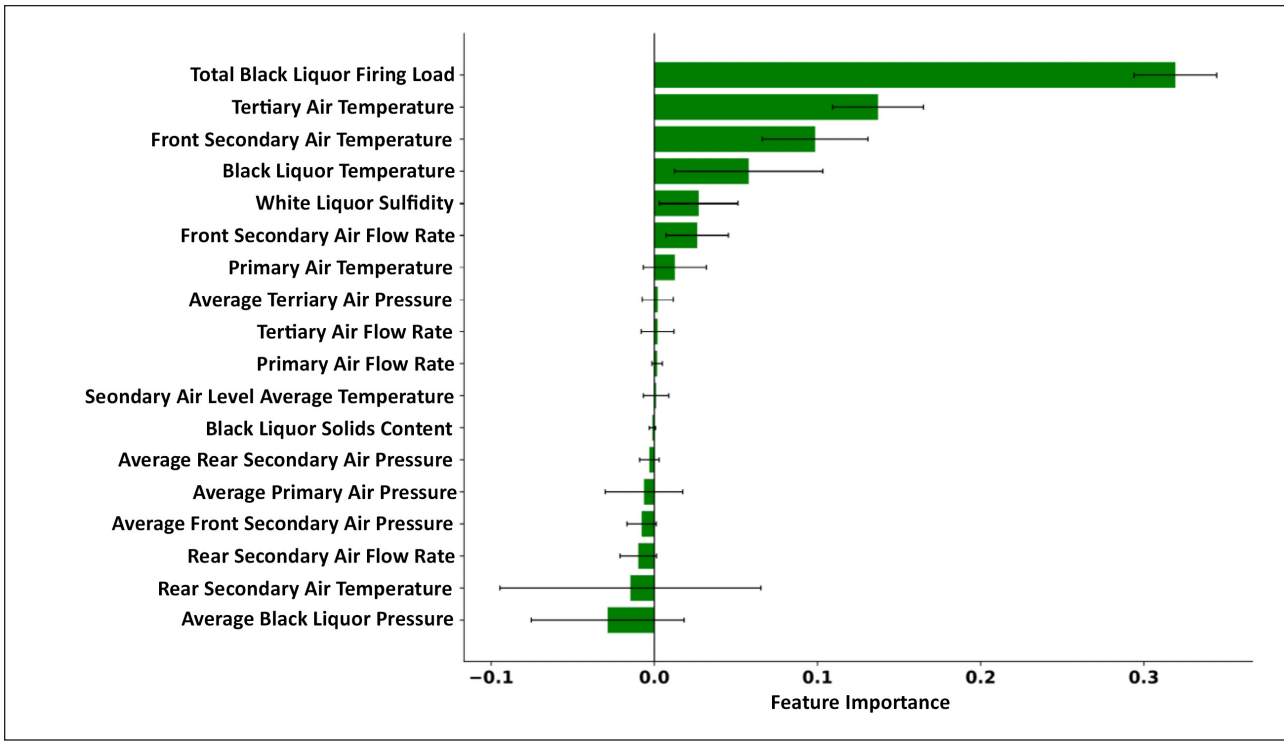
where m is the number of exogenous inputs, b_j is the coefficient associated with the jth exogenous input, and $n_{k,j}$ is the lag associated with the jth exogenous input. When the exogenous inputs are scaled to zero mean and unit variance, b_j represents the relative contribution of each exogenous input (**Fig. 6**).



4. Simulation outputs on test data (ΔP: flue gas pressure drop).



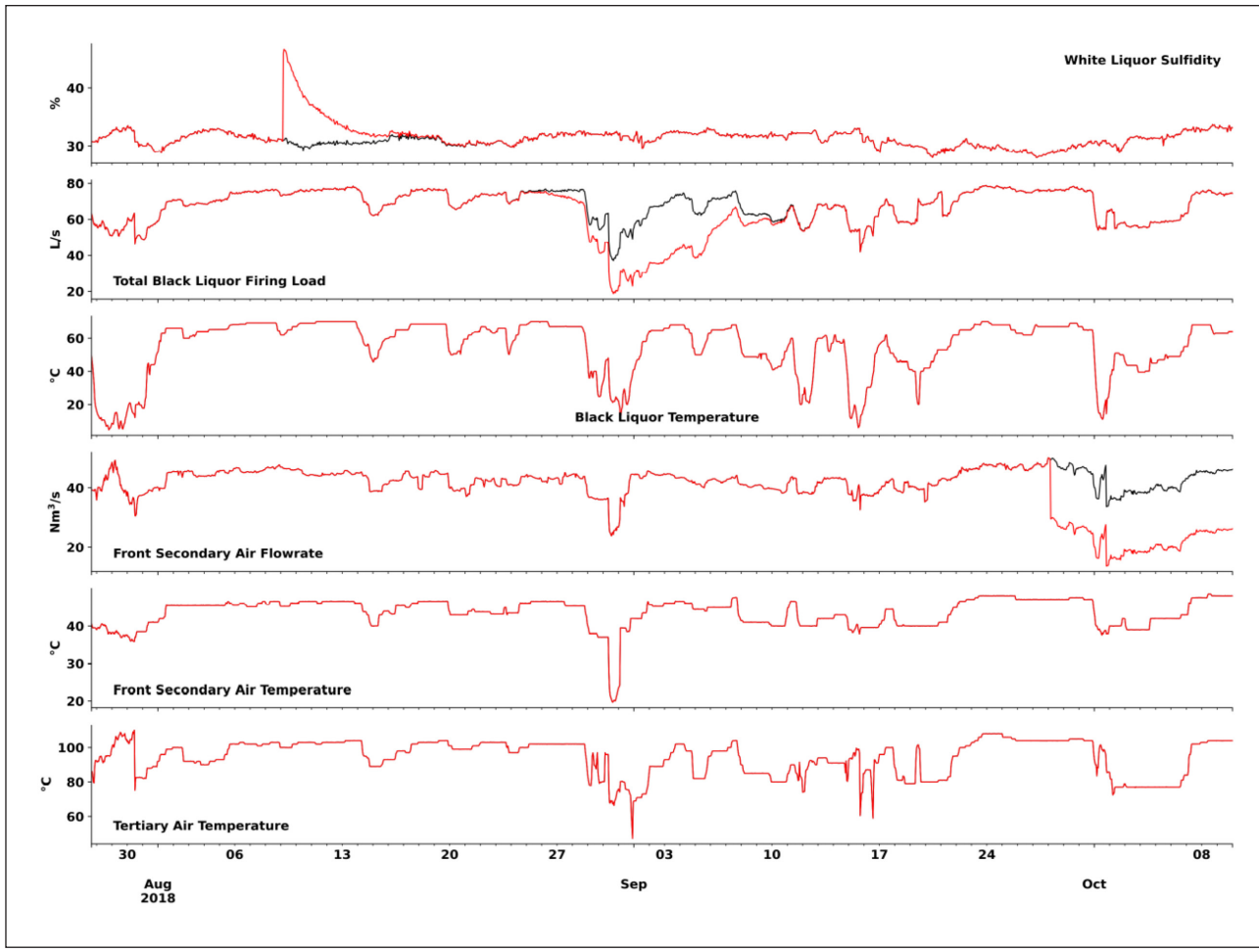
5. Parity plots and coefficient of determination (R^2) on unseen test data.



6. The autoregressive integrated exogenous (ARIX) model feature importance.

Of the 18 exogenous inputs, six were found to be significant and positively correlated with ΔP^*_i: black liquor firing load, tertiary air temperature, front secondary air temperature, black liquor temperature, white liquor sulfidity, and front secondary air flow rate. Importantly, these results do not suggest that the other inputs are irrelevant to boiler bank fouling. Rather, for this particular dataset and model, there is insufficient evidence to suggest that these other inputs are significant.

Another possibly relevant exogenous input is the number of liquor nozzles in operation. Since nozzle count was not used as an exogenous input, it is an unknown input whose effect may be accounted for by the noise term of the ARIX model. Including nozzle count may improve model



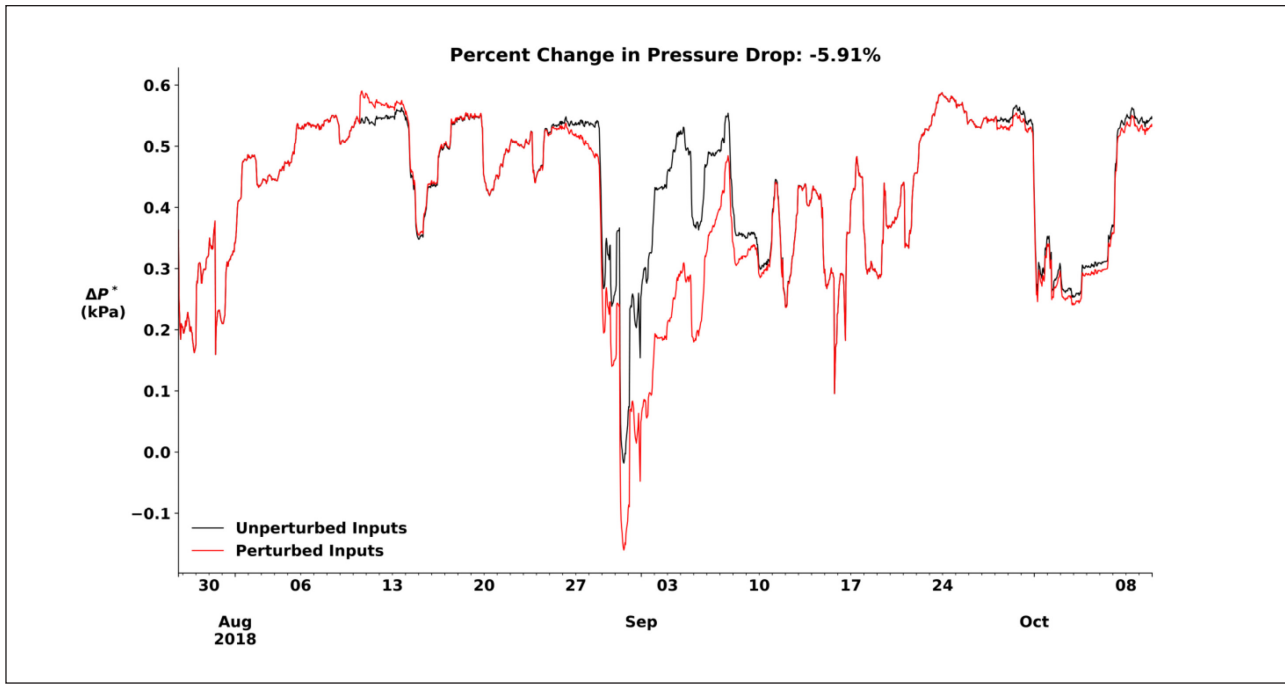
7. Perturbed inputs.

performance and decrease the variance of the identified noise term. This would be interesting to consider in a future investigation.

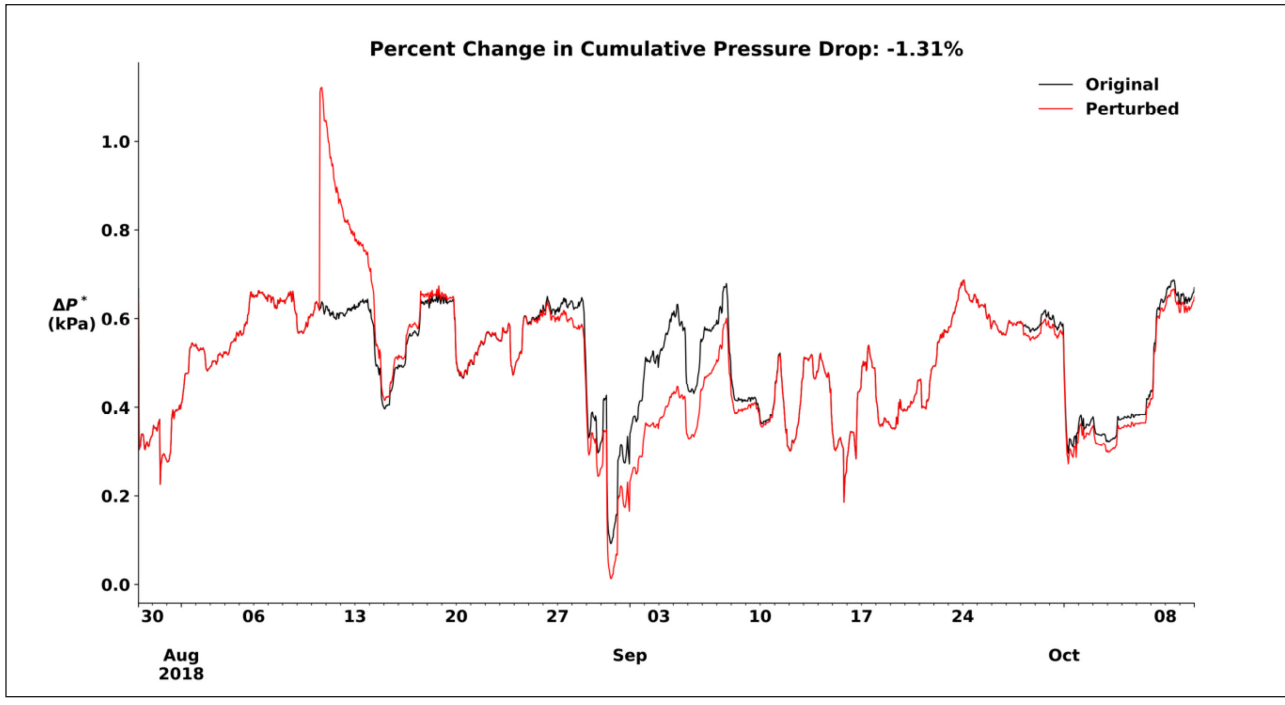
The identified, significant relationships are reasonable based on first-principles. Increasing the liquor firing load increases the number of droplets small enough to form carryover, increasing fouling. High flue gas temperatures are known to increase fouling, as higher air temperatures result in a higher volumetric flowrate of gas, increasing the likelihood of carryover. Liquor temperature affects the spray characteristics of black liquor fed to the boiler. Black liquor is fed to the boiler while under pressure and at temperatures above its boiling point. This is done to prevent flashing of liquor in the piping. In contrast, the boiler operates at slightly below atmospheric pressure. As a result, the liquor flashes upon entering the boiler. The degree of flashing increases with increasing liquor temperature, and a high degree of flashing creates significantly smaller droplets [38-39], which would lead to increased carryover. Increasing white liquor sulfidity increases the sulfidity of as-fired liquor, and increased as-fired liquor sulfidity is known to increase fireside fouling [3]. Higher sulfidity may result in fume that "glues" the carryover together, so a typical suggestion to reduce fouling is to reduce liquor sulfidity [1]. Importantly, changes in white liquor sulfidity would need to consider constraints allowed by digester operation. Additionally, the cost of less optimal digester performance may be offset by the savings of experiencing fewer boiler shutdowns due to excessive fouling. This balance should be determined on a mill-by-mill basis. A similar result was found by Versteeg and Tran [21]. Finally, increasing secondary air flow, which enters below the liquor guns, increases entrainment and carryover, increasing fouling.

Simulation of perturbed inputs

Since the identified ARIX model simulates a significant amount of the variability in ΔP^*_t, the model can be used to simulate ΔP^*_t given perturbations in mill parameters. To demonstrate, we injected three perturbations into the test data exogenous inputs (**Fig. 7**). First, we introduced a spike in white liquor sulfidity that gradually dissipates. The dynamics of this dissipation match an underlying noise model (an ARMA model) fit to the unperturbed white liquor sulfidity time series. Such a disturbance is called an innovational outlier [40]. Second, we introduced a decrease in total black liquor firing load. Initially, the rate of decrease accel-



8. The ARIX simulation.



9. Neural network simulation.

erates over time before reversing. This disturbance was introduced by superimposing an inverted normal distribution onto the unperturbed total black liquor firing load time series. Third, we introduced a sudden, constant decrease in front secondary air flow rate. This type of disturbance is called a level shift [41]. We perturbed different measurements in this manner to show how perturbations that vary in magnitude, dynamics, and duration can be simulated together. After perturbing the inputs, we passed the perturbed inputs through the ARIX model and compared the output to the output with unperturbed inputs (**Fig. 8**).

While, based on testing (Fig. 5), we are more confident in the ARIX model simulation of perturbed inputs, we include the neural network model simulation for demonstration (**Fig. 9**). The types of models considered in this work, trained on real process data to minimize simulation error,

allow for simulations of novel scenarios without subjecting the actual process to a potentially costly test. For example, for the studied dataset, a 0.5°C decrease in liquor firing temperature decreases the simulated ΔP^*_t by 0.12%. Another use of these models is deciding where to focus resources for process optimization. Various scenarios can be simulated to identify which inputs would, if optimized, provide the highest return on investment.

CONCLUSION

Here, we compared an ARIX model and a feedforward neural network model to simulate normalized boiler bank pressure drop ΔP^*_t, an indicator of boiler bank fouling. Both models can quantify a significant amount of variability in ΔP^*_t, and the ARIX model produces a more accurate simulation on unseen test data. The ARIX model parameters indicate that six of the 18 considered boiler parameters significantly affect ΔP^*_t, and their effects agree with first-principles. Finally, we showed that the trained models can be used to estimate how perturbations in boiler parameters would affect ΔP^*_t. **TJ**

ACKNOWLEDGEMENTS

This work is part of the industrial research consortium project "Effective Energy and Chemical Recovery in Pulp and Paper Mills – III" at the University of Toronto. Additionally, this work was supported by the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED

1. Vakkilainen, E.K., "Kraft Recovery Boilers — Principles and Practice," Finnish Recovery Boiler Committee, Helsinki, 2005.
2. Smook, G.A., *Handbook for Pulp & Paper Technologists*, 4th edn., TAPPI Press, Peachtree Corners, GA, USA, 2016.
3. Adams, T.N., *Kraft Recovery Boilers*, 2nd edn., TAPPI Press, Atlanta, 1997.
4. Wessel, R.A., Baxter, L.L., Shaddix, C., et al., "Particle formation and deposition in recovery boilers," *Int. Chem. Recovery Conf.*, TAPPI Press, 2004, p. 363.
5. Tran, H., "Recovery boiler fireside deposits and plugging prevention," *TAPPI Kraft Recovery Course*, TAPPI Press, Atlanta, 2007, pp. 537-572.
6. Tran, H., *Kraft Recovery Boilers*, 3rd edn., TAPPI Press, Peachtree Corners, GA, USA, 2018.
7. Mueller, C., Brink, A., and Hupa, M., *Comput. Aided Chem. Eng.* 14: 809(2003). https://doi.org/10.1016/S1570-7946(03)80216-X.
8. Box, G., and Narasimhan, S., *Qual. Eng.* 22(2): 60(2010). https://doi.org/10.1080/08982110903510297.
9. Aström, K.J., *IBM J. Res. Dev.* 11(4): 389(1967). https://doi.org/10.1147/rd.114.0389.
10. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., et al., *Time Series Analysis: Forecasting and Control*, 5th edn., John Wiley & Sons, Hoboken, NJ, USA, 2015.
11. Ljung, L., *System Identification: Theory for the User*, 2nd edn., Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
12. Gopaluni, R.B., Tulsyan, A., Chachuat, B., et al., *IFAC-PapersOnLine* 53(2): 218(2020). https://doi.org/10.1016/j.ifacol.2020.12.126.
13. Ihme, M., Chung, W.T., and Mishra, A.A., *Prog. Energy Combust. Sci.* 91: 101010(2022). https://doi.org/10.1016/j.pecs.2022.101010.
14. Liao, M. and Yao, Y., *GCB Bioenergy* 13(5): 774(2021). https://doi.org/10.1111/gcbb.12816.
15. Mohammadi, K., Immonen, J., Blackburn, L.D., et al., *Rev. Chem. Eng.* 39(6): 1027(2023). https://doi.org/10.1515/revce-2021-0107.
16. Sitapure, N. and Kwon, J.S.-I., *Chem. Eng. Res. Des.* 194: 461(2023). https://doi.org/10.1016/j.cherd.2023.04.028.
17. Almeida, G.M., Park, S.W., and Cardoso, M., *IFAC Proc. Vols.* 37(16): 91(2004). https://doi.org/10.1016/S1474-6670(17)30856-X.
18. Jones, A., Wagoner, J., Michaelson, T., et al., "Use of multivariate analysis to understand the root causes of premature smelt spout failures on recovery boilers," *Int. Chem. Recovery Conf.*, PAPTAC, Montreal, 2014, p. 10.
19. Pereira Parente, A., de Souza Jr., M.B., Valdman, A., et al., *Processes* 7(12): 958(2019). https://doi.org/10.3390/pr7120958.
20. Sainlez, M. and Heyen, G., *J. Comput. Appl. Math.* 246: 329(2013). https://doi.org/10.1016/j.cam.2012.06.026.
21. Versteeg, P. and Tran, H., *TAPPI J.* 8(11): 22(2009). https://doi.org/10.32964/TJ8.11.22.
22. Belisario, A.B., Edberg, A., Bjork, M., et al., *TAPPI J.* 22(3): 162(2023). https://doi.org/10.32964/TJ22.3.162.
23. Box, G.E.P., Luceño, A., and del Carmen Paniagua-Quinones, M., *Statistical Control by Monitoring and Adjustment*, 2nd edn., John Wiley & Sons, Hoboken, NJ, USA, 2011.
24. del Castillo, E., *Statistical Process Adjustment for Quality Control*, John Wiley & Sons, New York, 2002.
25. Romeo, L.M. and Gareta, R., *Biomass Bioenergy* 33(5): 854(2009). https://doi.org/10.1016/j.biombioe.2009.01.008.
26. Romeo, L.M. and Gareta, R., *Eng. Appl. Artif. Intell.* 19(8): 915(2006). https://doi.org/10.1016/j.engappai.2006.01.019.
27. Romeo, L.M. and Gareta, R., *Appl. Therm. Eng.* 26(14): 1530(2006). https://doi.org/10.1016/j.applthermaleng.2005.12.006.
28. Audibert, J., Michiardi, P., Guyard, F., et al., *Pattern Recognit.* 132: 108945(2022). https://doi.org/10.1016/j.patcog.2022.108945.
29. Hansen, H.H., Kulahci, M., and Nielsen, B.F., *Comput. Chem. Eng.* 178: 108391(2023). https://doi.org/10.1016/j.compchemeng.2023.108391.
30. Li, J. and Qin, S.J., *Comput. Chem. Eng.* 175: 108264(2023). https://doi.org/10.1016/j.compchemeng.2023.108264.
31. Rewicki, F., Denzler, J., and Niebling, J., *Appl. Sci.* 13(3): 1778(2023). https://doi.org/10.3390/app13031778.
32. Wu, R. and Keogh, E.J., *IEEE Trans. Knowl. Data Eng.* 35(3): 2421(2023). https://doi.org/10.1109/TKDE.2021.3112126.
33. Legaard, C., Schranz, T., Schweiger, G., et al., *ACM Comput. Surv.* 55(11): 1(2023). https://doi.org/10.1145/3567591.
34. Chinvornrat, S., Lu, B. Huang, J.-K., et al., *Proc. 1999 Am. Control Conf.* (Cat. No. 99CH36251) 1: 331(1999). https://doi.org/10.1109/ACC.1999.782795.

35. Vakkilainen, E., "Troubleshooting recovery boiler fouling," *Conf. Behavior Inorganic Material Recovery Boilers*, American Institute of Chemical Engineers, New York, 2000.

36. Vakkilainen, E. and Vihavainen, E., "Long-term fouling of recovery boiler surfaces," *Int. Chem. Recovery Conf.*, TAPPI Press, Atlanta, 1992.

37. Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016.

38. Bennington, C.P.J., "The effect of temperature on drop size of black liquor sprays," University of British Columbia, MASc thesis, 1983. https://doi.org/10.14288/1.0058907.

39. Miikkulainen, P., "Spray formation of high dry solids black liquor in recovery boiler furnaces," Ph.D. dissertation, Helsinki University of Technology, Helsinki, 2006.

40. Tsay, R.S., *J. Forecast.* 7(1): 1(1988). https://doi.org/10.1002/for.3980070102.

41. Box, G.E.P. and Tiao, G.C., *Biometrika* 52(1/2): 181(1965). https://doi.org/10.2307/2333823.

ABOUT THIS PAPER

Cite this article as:
Ng, J., De Almeida, G.M., Vakkilainen, E.K., et al., *TAPPI J.* 23(7): 374(2024). https://doi.org/10.32964/TJ23.7.374

DOI: https://doi.org/10.32964/TJ23.7.374

ISSN: 0734-1415

Publisher: TAPPI Press

About this journal

ABOUT THE AUTHORS

Recovery boiler fouling bottlenecks production and limits thermal efficiency. Given the difficulties in modeling boiler fouling, we wanted to compare classical techniques against deep learning, which has become increasingly popular. While other studies have explored deep learning for modeling mill processes, benchmarking against simpler methods is usually poor or ignored. In this study, the benchmarking is more rigorous, and we account for autocorrelation by distinguishing between prediction and simulation, which is also typically ignored.

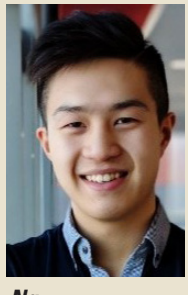
The most difficult aspect of this study was training the neural network to minimize simulation error. In the literature, neural networks are usually trained to minimize prediction error, so most deep learning frameworks are conveniently designed for such. In our case, we had to customize the training procedure.

Before this research, we considered prediction and simulation to be synonymous. Instead, this research showed that prediction and simulation are distinct and that this distinction affects the necessary complexity and applicability of the resulting model.

The similar performances between the linear model and the neural network were surprising. Originally, we had expected the increased complexity of the neural network to result in a more accurate model.

This study shows that deep learning and other machine learning techniques are not inherently beneficial for modeling mill phenomena. Given our results, mills should seriously consider simpler methods before resorting to more sophisticated ones. Further, our discussions of stochasticity and autocorrelation can help mills clarify the goals of a model. The next step in our research would be to apply these models for control, with adjustments based on model predictions.

Ng is a Ph.D. candidate, Laryshyn is professor, and DeMartini is associate professor and director of the Pulp and Paper Centre at the University of Toronto in Toronto, ON, Canada. De Almeida is researcher and Vakkilainen is professor at LUT University in Lappeenranta, Finland. Email Ng at jerrymh.ng@mail.utoronto.ca.


Ng


De Almeida


Vakkilainen


Laryshyn


DeMartini