

Chemical Pulping

Jerry Ng*, Yuri Lawryshyn and Nikolai DeMartini

Estimating lags in a kraft mill

<https://doi.org/10.1515/npprj-2024-0004>

Received January 15, 2024; accepted April 26, 2024;

published online May 17, 2024

Abstract: In pulp mills, lags obscure the effect of upstream operations on downstream measurements. Here, we estimate lags in a Canadian pulp mill using autoregressive exogenous (ARX) models. First, we show that ARX models can approximate lags in a process simulation that resembles the liquor storage tanks in pulp mills, a major source of lag in the kraft recovery cycle. Then, we use ARX models to estimate the lagged effect of a change in species pulped on as-fired liquor heating value, viscosity, and boiling point rise. Additionally, we compare the predictions of the ARX models to autoregressive (AR) models and a persistence model. The estimated lags between a change in species and heating value (49 h) and boiling point rise (41 h) agree with a detailed simulation of the mill and are close to estimated hydraulic residence times, suggesting that the liquor tanks exhibit imperfect mixing. A lagged effect of species change on viscosity could not be identified. ARX and AR models produce similar predictions that are slightly better than those of a persistence model. Finally, we show that process measurements upstream of units characterized by large residence times will likely provide little benefit to prediction accuracy.

Keywords: kraft pulping; system identification; black liquor; time series analysis; machine learning

1 Introduction

Pulp mills comprise storage tanks that delay the impacts of upstream disturbances on downstream operations. This

is helpful in preventing issues in one area of a mill from spreading to other mill areas. However, as a result, models relating measurements upstream and downstream of these storage tanks must account for such delays. Here, we define lag as the time before a change upstream begins to significantly affect a downstream measurement. While, in some cases, lags are approximately equal to hydraulic residence times, this need not be the case. Mixing, other transport phenomena, reaction kinetics, and non-idealities can cause the time before significant impact (the lag) and the hydraulic residence time to differ (Ljung 1998). Lags are particularly important when developing machine-learning models or applying statistical techniques as lags affect how data should be pre-processed before model fitting. Ideally, these lags would be identifiable from normal operating data alone (Box et al. 2015).

Black liquor, an aqueous by-product of the pulping process, is concentrated before being burned in a recovery boiler to recover pulping chemicals and energy. The liquor fed into the boiler is as-fired liquor. Specifically, our goal is to estimate the lag between a change in species pulped and as-fired heating value, viscosity, and boiling point rise (BPR). Heating value, typically reported as higher heating value (HHV), is important as the heat of combustion drives the recovery of sodium sulfide and production of steam in the recovery boiler (Adams 1997). Viscosity and BPR affect droplet formation as liquor is sprayed into the recovery boiler which, if poorly controlled, can result in excess carryover, fouling, and depletion of the char bed (Miikkulainen et al. 2005). All three liquor properties are known to vary with wood species (Hart 2022). While lag estimation in mills and other process data has been investigated elsewhere, to our knowledge, in a mill, the lag between a change in species pulped and the as-fired properties has not been explored.

1.1 Objectives

In this study, our objective is to identify the lag between changes in wood species pulped and the corresponding changes in as-fired black liquor HHV, viscosity, and BPR based only on process data. If these lags could be identified, operators could better anticipate when process parameters would be affected by upstream disturbances, allowing them to act preemptively to avert negative impacts. Further,

*Corresponding author: Jerry Ng, Chemical Engineering and Applied Chemistry, University of Toronto Faculty of Applied Science & Engineering, Toronto, Ontario, Canada, E-mail: jerrymh.ng@mail.utoronto.ca.
<https://orcid.org/0000-0002-8547-2642>

Yuri Lawryshyn, Chemical Engineering and Applied Chemistry, University of Toronto Faculty of Applied Science & Engineering, Toronto, Ontario, Canada

Nikolai DeMartini, University of Toronto Faculty of Applied Science and Engineering, Toronto, Ontario, Canada

by using a data-driven approach instead of heuristics, lag estimates would be repeatable, systematic, and not rely solely on process experience.

2 Process lag estimation in the literature

Multiple studies in the recent literature investigate process lag estimation. Generally, the approaches can be divided into six categories (Table 1). In many mills and in some studies (Correia et al. 2018), lags are estimated heuristically. While these estimates can be accurate, they rely on process experience which can take years to acquire. In contrast, we focus on data-driven methods that do not require as much process experience. In this section, we discuss data-driven process lag estimation approaches in the recent literature along with their limitations.

2.1 Hydraulic residence time

One approach is to assume that hydraulic residence times are equal to process lags. A simple hydraulic residence time can be used:

Table 1: Categorization of lag estimation methods in the recent literature.

Lag estimation method	Example
Heuristic	Correia et al. (2018)
Hydraulic residence time (Zenger and Ylinen 1994)	Sayda and Taylor (2003)
Lipschitz index (He and Asada 1993)	Wang and Chen (2006) Sainlez and Heyen (2013) Rebello et al. (2022)
Correlation analysis	Dahlbäck (2020) Graziani and Xibilia (2021) Topalian et al. (2022) Ekbåge et al. (2019)
Feature engineering	Lee et al. (2006) Galicía et al. (2011) Graziani and Xibilia (2019) Graziani and Xibilia (2021) Chen and Zhao (2020) John and Ferbinteanu (2021)
Hyperparameter tuning	Ettaleb et al. (1998) Almeida and Park (2013) Matos et al. (2008) Mechaqrane and Zouak (2004) Chetouani (2008) El-Din and Smith (2002) Galicía et al. (2011)

$$T_d(t) = \frac{V(t)}{Q(t)}, \quad (1)$$

where T_d is residence time, V is tank volume, and Q is the volumetric flowrate through the tank. This approach is the simplest, quickly providing a rough lag estimate. Zenger and Ylinen (1994) expanded on this residence time calculation:

$$V(t) = \int_{t-T_d(t)}^t Q(\tau) d\tau. \quad (2)$$

Unlike Equations (1) and (2) accounts for variations in volumetric flowrate and tank volume as fluid travels through the tank. Sayda and Taylor (2003) used Equation (2) to estimate lags in a mechanical pulp bleaching process. Unfortunately, as mentioned earlier, hydraulic residence times may incorrectly predict lags as a result of ignoring mixing effects, reaction kinetics, and non-idealities.

2.2 Lipschitz index

In order to account for non-idealities, a popular alternative is the method by He and Asada (1993). They define the Lipschitz quotient

$$q_{i,j} = \frac{|y_i - y_j|}{|x_i - x_j|}, \quad i \neq j, \quad (3)$$

where y is the system output, x is the system input. This can be expanded to

$$q_{i,j}^n = \frac{|\delta y|}{\sqrt{(\delta x_1)^2 + \dots + (\delta x_n)^2}} \quad (4)$$

and used to calculate the Lipschitz index q of model order n

$$q^n = \left(\prod_{k=1}^p \sqrt[n]{q^n(k)} \right)^{\frac{1}{p}}, \quad (5)$$

where $q^n(k)$ is the k th largest Lipschitz quotient among all $q_{i,j}^n$, and p is a positive number (usually $p = 0.01N$). Generally, the Lipschitz index q^n decreases as model order n increases. The value of n at which q^n plateaus is the chosen model order. This approach has been used in a number of studies (Rebello et al. 2022; Sainlez and Heyen 2013; Wang and Chen 2006). It is popular because it has been shown to identify lags in non-linear systems. Unfortunately, this method can result in prohibitively large input sizes, and if the lags are large or if the sampling frequency is high, uninformative input data will be retained, bloating the model. The importance of minimizing model size is discussed by Galicía et al. (2011).

2.3 Hyperparameter tuning

Other approaches treat lag estimation as a type of hyperparameter tuning. Souza and Araújo (2011) combined forward-selection with a support vector machine model, treating increasingly lagged values of inputs as possible, additional inputs. Their approach suffers the same drawback as the Lipschitz index: uninformative inputs can be retained. Almeida and Park (2013) used trial-and-error to estimate the necessary time delay for a neural network model of an evaporation plant in a pulp mill. However, this brute force method can become intractable if the model is large and if many possible time delays must be checked.

2.4 Correlation analysis

Sometimes, the cross-correlation between the output and varying lagged values of the input are compared to identify lags. Graziani and Xibilia (2019) estimated lags based on the cross-correlation of engineered features, going as far as to experiment with other types of cross-correlation metrics (Graziani and Xibilia 2021). However, the proposed correlations do not account for autocorrelation in the input (Bisgaard and Kulahci 2011; Box et al. 2015). Since the inputs are usually autocorrelated, their impact on the output is “smeared” over time. By not accounting for the dynamics of the input, spurious correlations emerge, such as positive cross-correlations between input measurements in the future and output measurements in the past.

2.5 Feature engineering

Finally, some studies transform process data into a latent space and estimate the lags in the latent space. For example, Chen and Zhao (2020) used dynamic principal component analysis (diPCA) to transform process data into latent variables and compared the cross-correlations between the latent variables and other variables, where peak cross-correlations indicated lags. Similarly, John and Ferbinteanu (2021) transformed time series data into graphs before performing cross-correlation analysis. Unfortunately, due to the use of engineered features, these methods can be difficult to interpret.

2.6 ARX models

In this study, we model the relationships for a variety of input-output pairs in a simulated two-tank system and in a

Canadian kraft pulp mill using autoregressive exogenous (ARX) models, which account for autocorrelations in both the input and output. In the process of identifying these models, the lags between inputs and outputs are identified. Ettaleb et al. (1998) demonstrated this approach with simulated autoregressive moving-average exogenous (ARMAX) processes, and Matos et al. (2008) used ARX models to model the relationship between digester parameters and the viscosity of bleached pulp. While Galicia et al. (2011) used DPLS regression to approximate kappa number, they considered using ARX models to estimate the necessary amount of historical data and dead times to incorporate into the model. Wasson (2016) used Box-Jenkins transfer function models, which belong to the same family of models as ARX models (Ljung 1998), to estimate the lags of inputs to a hog fuel boiler.

We first validate the use of ARX models using a simple, numerical simulation of two tanks in series with an added artificial delay. Then, we apply the approach to mill data. The estimated lags of the mill are compared to estimated hydraulic residence times and the estimated lags from a detailed simulation of the mill (Rogerson 2022).

3 Methodology

3.1 Two tank simulation

Liquor storage tanks are predicted to create the largest lags in mills. As a result, using CADSIM Plus (Aurel Systems 2019), we developed a simulation of a water-NaCl solution flowing through two tanks in series (Figure 1). In the simulation, tank volumes, flowrates, and feed concentration are manipulated. Proportional-integral level controllers with noisy, oscillating setpoints and large integral terms manipulate flowrates to simulate the oscillating tank level control in a mill: while a mill pulps softwood, the volume of stored black liquor increases; during hardwood operation, the volume of stored liquor decreases. Additionally, all simulation parameters can be injected with noise. The magnitude of this noise varies for each parameter and is indicated in Table 2. Finally, a transport lag (box labelled “delay”) is placed before the first tank to simulate imperfect mixing (i.e., a dead-time). Once the simulation was initiated (Table 2), it was allowed to run for 15,000 time steps (1 time step = 1 min). The first 2000 time steps of data were not analyzed as the simulation was reaching steady-state during this time.

The simulation data is shown in Figure 2. All measurements exhibit noise except for the outlet flowrate from Tank 2, which was oscillated without noise. While the feed concentration is quite noisy, most of that noise is attenuated in Tank 1. This is likely due to the large volume of fluid in Tank 1 (initially 250 m³) relative to its inlet flowrate (30 LPS): the NaCl is diluted in the tank. The marginal amount of noise in the concentration exiting Tank 1 is attenuated in Tank 2. This is expected: the inlet flowrate of Tank 2 is similar to Tank 1, but Tank 2 contains less fluid to dilute its inlet (50 m³ vs. 250 m³).

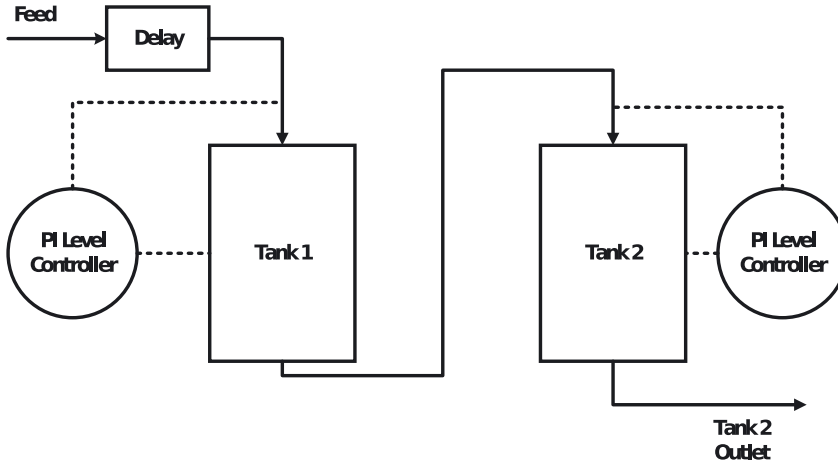


Figure 1: Schematic of the two-tank simulation.

Table 2: Two-tank simulation initial conditions.

Parameter	Initial value	Noise primary standard deviation	Noise secondary standard deviation	Oscillation amplitude	Oscillation frequency	Oscillation noise primary standard deviation	Oscillation noise secondary standard deviation	Ramp rate
Tank 1 volume	500 m ³	None	None	None	None	None	None	None
Tank 2 volume	100 m ³	None	None	None	None	None	None	None
Feed water flowrate	30 LPS	None	None	None	None	None	None	None
Feed NaCl flowrate	0.1 kg/s	0.05	0.01	None	None	None	None	None
Tank 1 level SP	50 %	5	1	10	0.01	3	1	1
Tank 2 level SP	50 %	5	1	10	0.001	3	1	1
Tank 2 outlet flowrate	20 kg/s	None	None	5	0.01	None	None	0.1
Transport lag	10 min	None	None	None	None	None	None	None

3.2 Mill data

Mill data (Figure 3) was obtained from a North American mill that pulps both hardwood and softwood. The dataset spans November 2022 to April 2023, contains 90 measurements from the digester to the recovery boiler, and was sampled at a 10-min frequency. The 90 measurements were chosen based on mill expertise, and the data was filtered for normal operation: digester production rate above 400 ADMTPD and recovery boiler production rate above 1.5 tons per day DS. HHV was calculated as

$$\text{HHV} = \frac{A+20}{0.7} \frac{y}{24} - \frac{g}{24} r \quad (6)$$

where A is the total flowrate of air to the boiler, y is an excess air characterization, g is the flowrate of air consumed by natural gas, and r is the liquor firing rate. The mill verified Equation (6) against laboratory measurements of black liquor heating value. Viscosity was calculated using

$$\text{Viscosity} = 0.1351L^3 - 26.8456L^2 + 1795.92L - 40314.5, \quad (7)$$

where L is the pump motor load. Equation (7) was developed in-house by the mill using an online viscometer which measured a sidestream of the as-fired tank's re-circulation pump. Finally, BPR was calculated as

$$\text{BPR} = T_{\text{Liq}} - T_{\text{Vap}}, \quad (8)$$

where T_{Liq} is the temperature of liquor exiting the as-fired tank and T_{Vap} is the temperature of the vapour exiting the as-fired tank. The temperature of the vapour is estimated based on the pressure of the vapour and a steam table, assuming that the vapour exiting the tank is at saturation.

Equations (6) and (7) are empirical correlations while Equation (8) is the definition of BPR. Measurements falling outside of the 0.5 and 99.5 percentiles were dropped to remove spurious measurements, and missing data (roughly 12 % of the data) was imputed using linear interpolation.

3.3 ARX models

ARX models, which belong to the family of polynomial models described by Ljung (1998), represent the output as a linear combination of lagged inputs and outputs:

$$\begin{aligned} A(q)Y_t &= B(q)X_t + \varepsilon_t, \\ A(q) &= 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \\ B(q) &= b_1q^{-n_k} + \dots + b_{n_b}q^{-n_k-n_b}, \\ Y_t &= a_1Y_{t-1} + \dots + a_{n_a}Y_{t-n_a} + b_1X_{t-n_k} + \dots + b_{n_b}X_{t-n_k-n_b} + \varepsilon_t, \end{aligned} \quad (9)$$

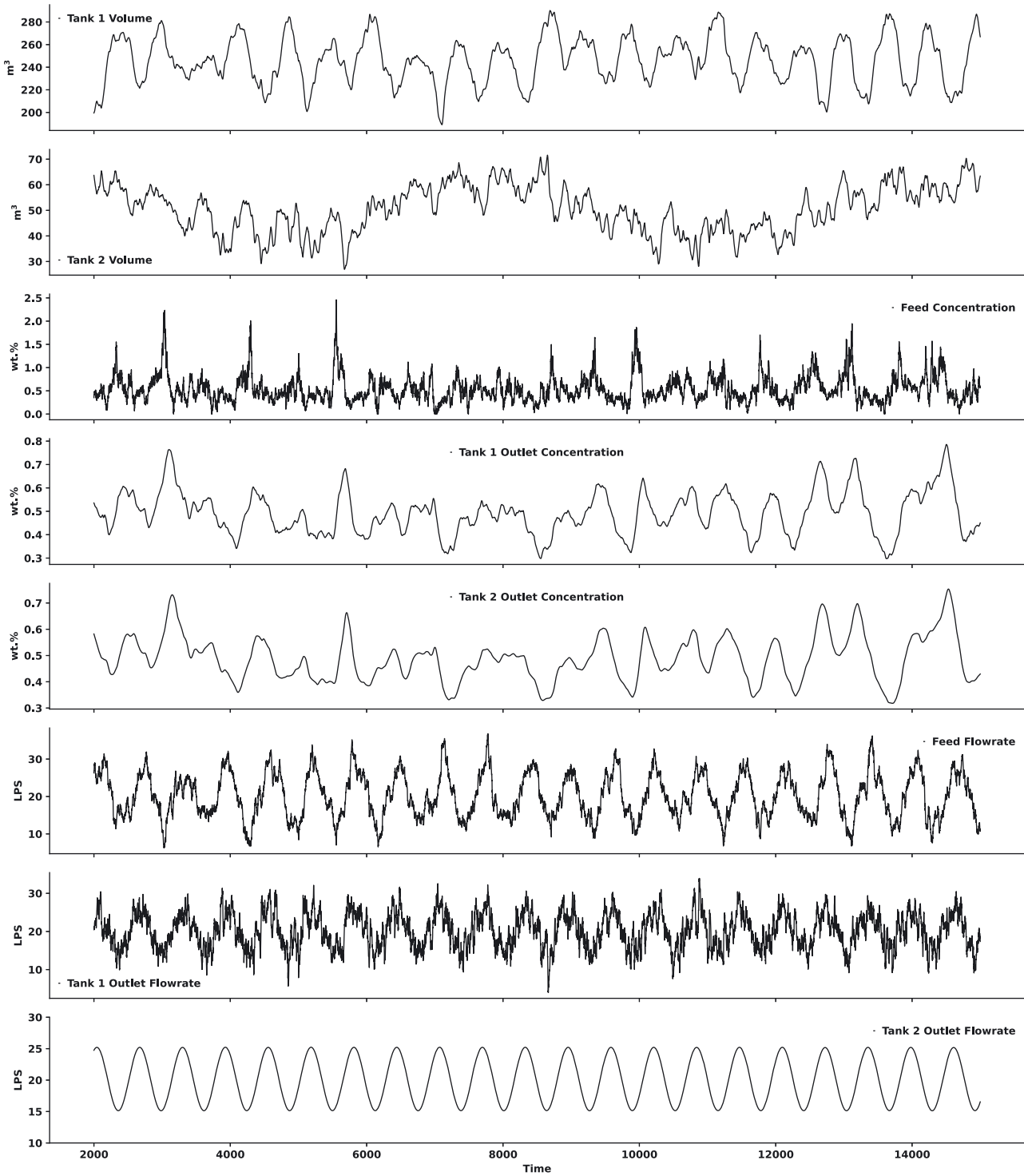


Figure 2: Data from two-tank simulation.

where Y_t is the system output; X_t is the system input; n_a and n_b are the lag orders of the polynomials $A(q)$ and $B(q)$, respectively; q is the shift operator; and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$ is a white noise process. Importantly, n_k is the lag between input and output, and its identification is the focus of this

work. Equation (9) is analogous to the combination of a filtered input and filtered white noise. This is the preferred representation by Ljung (1998) and Box et al. (2015) as it makes explicit that the system is driven by noise ε_t .

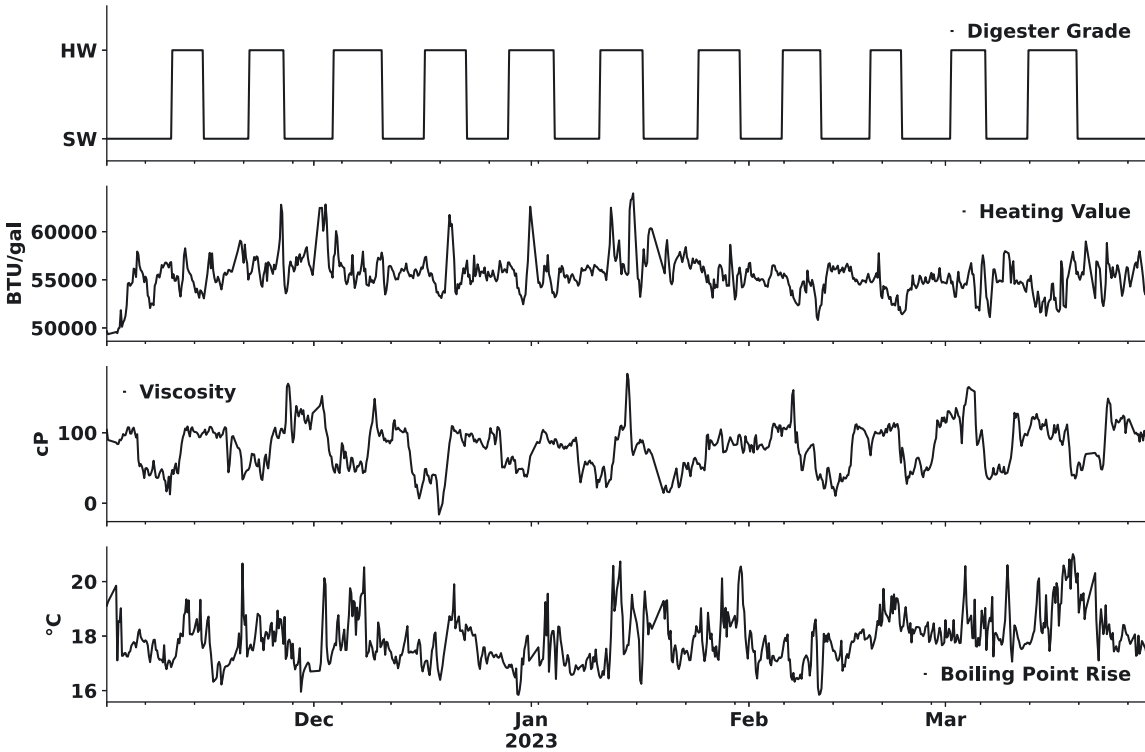


Figure 3: Species changes and as-fired liquor properties (4 h averaged).

3.4 Model fitting procedure

The same fitting procedure was used for both the simulation data and the mill data. Model identification was done using the System Identification Toolbox in MATLAB 2022 (The MathWorks Inc 2022). First, the data was split into training, validation, and testing sets. For the simulation data, the first 4000 rows were used as training data, and the proceeding 2000 rows were used as validation data, resulting in a 30/15/55 train/validation/test split. For the mill data, the first 1500 rows were used as training data, and the proceeding 500 rows were used as validation data, resulting in a 28/15/57 train/validation/test split. For fitting linear models, datasets on the order of hundreds of samples are typical. We favoured a smaller training and validation set in order to create as large of a testing set as possible. A larger unseen testing set leads to a more robust assessment of the fitted model and estimated lags. Additionally, smaller training and validation sets demonstrate that, in practice, lags can be estimated sooner: less time is spent waiting for sufficient data.

After partitioning the data, ARX models with $n_a = [1:10]$, $n_b = [0:10]$, and $n_k = [0:50]$ were evaluated. The possible orders of lagged inputs n_b include 0, allowing for the input to be ignored if beneficial. For each input-output pair, the selected model was the one that maximized the explained variance in the validation data. The coefficients of the identified ARX models are shown in Table 3. The performances of the identified ARX models against the testing data were compared to persistence models (no fitting procedure required) and autoregressive (AR) models. AR models were fitted in a similar manner as ARX models, except digester grade was not considered ($B(q) = 0$). The persistence model is an ARX model, where $A(q) = 1 - q^{-1}$ and $B(q) = 0$.

Table 3: ARX model coefficients.

HHV	$A(q)$	$1 - 0.6164q^{-1} - 0.3835q^{-2}$
	$B(q)$	$-158.1q^{-49} - 41.27q^{-50} - 273.1q^{-51} + 722.8q^{-52} - 18.54q^{-53} - 11.95q^{-54} - 493.9q^{-55} + 641.5q^{-56} - 367.3q^{-57}$
Viscosity ^a	$A(q)$	$1 - 0.9917q^{-1}$
	$B(q)$	$4.096q^{-11} - 6.529q^{-12} + 3.057q^{-13} - 2.095q^{-14} + 1.982q^{-15} - 0.4065q^{-16} - 2.529q^{-17} + 4.124z^{-18} - 4.308z^{-19} + 3.651z^{-20}$
BPR	$A(q)$	$1 - 1.068q^{-1} + 0.1896q^{-2} - 0.02325q^{-3} - 0.09807q^{-4}$
	$B(q)$	$-0.1079q^{-41} + 0.1457q^{-42} - 0.03551q^{-43} + 0.1035q^{-44} - 0.09997q^{-45} + 0.1846q^{-46} - 0.2691q^{-47} - 0.003473q^{-48} - 0.1942q^{-49} + 0.2724q^{-50}$

^aDoes not outperform AR model: lag likely incorrect.

3.5 Model performance metrics

Model performances were evaluated based on three metrics: coefficient of determination (R^2), mean absolute percentage error (MAPE), and root mean squared error (RMSE). If over n samples, y is the true value, \hat{y} is the predicted value, and \bar{y} is the average of true values, then R^2 , MAPE, and RMSE are calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (10)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|}, \quad (11)$$

and

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (12)$$

respectively.

4 Results and discussion

4.1 Two-tank simulation

With the two tank simulation data, the hydraulic residence times calculated using Equation (2) overestimate the true lags whereas the lag estimates of the ARX models are close to the true lags (Table 4). This was expected as Equation (2) assumes a plug flow through vessels while the simulated tanks are perfectly mixed. These results show that ARX models can identify lags that differ from hydraulic residence times.

4.2 Mill data

While a lagged effect between digester grade and as-fired viscosity is weak and likely insignificant, the estimated lags of the ARX models relating digester grade to HHV and BPR seem reasonable (Table 5) and improve predictions (Table 7). The estimated lags agree with those proposed by Rogerson (2022) who simulated the same mill with constant tank volumes. Rogerson estimated that, when transitioning from softwood to hardwood, the initial presence of hardwood organics in the as-fired liquor occurs 15 h after species change and that the as-fired liquor is composed of 95 % hard wood organics after 65 h. Similarly, when transitioning from hardwood to softwood, the initial and 95 % changes are seen at 12 h and 41 h, respectively. The ARX models were identified using data that contained both types of transitions, so it is reasonable that the lag that optimizes model performance is slightly longer than that estimated for hardwood to softwood transitions and slightly shorter than that estimated for softwood to hardwood transitions. The lag

between digester grade change and BPR is expected to be lower as the BPR is calculated at the final storage tank, which is separated from the as-fired liquor ring header by the volume of the storage tank. We estimated the hydraulic residence time of this tank to be 4–5 h based on Equation (2), and Rogerson (2022) estimates this lag to be roughly 7 h.

Additionally, we combined our estimated hydraulic residence times of storage tanks with the residence time estimates of smaller unit operations by Rogerson (2022) to obtain total hydraulic residence time estimates between a change in digester grade and properties of as-fired liquor (Table 5). Except for as-fired viscosity, where the estimated lagged effect is weak and likely incorrect, the estimated hydraulic residence times are close to the ARX model lag estimates.

Unlike the two tank simulation results, with the mill data, the hydraulic residence times and estimated lags are close. This suggests that the liquor storage tanks exhibit imperfect mixing. Again, Equation (2) assumes a plug flow through the vessel. If there were significant mixing in any of the storage tanks throughout the process, the ARX model lag estimates would be lower than the hydraulic residence time estimates, like in the two-tank simulation. Instead, the lag estimates are close to the hydraulic residence time estimates. Likely, the volume of liquid inside the mill storage tanks are so large relative to the volumetric flowrate passing through that the liquor inside the tank is minimally agitated. As a result, liquor fed to the tanks is poorly mixed and lingers.

4.3 Comparing predictions of ARX and AR models

While a lag may be identified during the fitting of an ARX model ($n_k \neq 0$), a final check using simpler models and unseen test data is important to detect spurious estimates. In this study, the ARX models are bivariate: they consider both past output measurements and an exogenous input

Table 4: Lag estimates for two tank simulation.

Input – output	Actual lag (min)	Median hydraulic residence time estimate (min)	ARX model lag estimate (min)
Feed – tank 1 outlet	10	203	10
Feed – tank 2 outlet	10	245	11
Tank 1 outlet – tank 2 outlet	0	42	1

Table 5: Lag estimates for mill data.

Input – output	Median hydraulic residence time estimate (h)	ARX model lag estimate (h)
Digester grade – black liquor HHV	43	49
Digester grade – black liquor viscosity	43	11 ^a
Digester grade – black liquor BPR	38	41

^aLikely insignificant.

(species). If, on unseen test data, the ARX model predictions are more accurate than a model that does not consider species, i.e., an AR model, then there is evidence that the estimated effect of species, including its lag, is significant. Similarly, if an AR model fails to outperform a simple persistence model, that suggests the autocorrelation identified by the AR model is insignificant. In this section, we compare the accuracies of the persistence model, the fitted AR models, and the fitted ARX models in both the two-tank simulation and the mill data.

For the two tank simulation, ARX models provide a small improvement in all metrics over AR models, which are more accurate than a persistence model (Table 6). Similarly, for the mill HHV data, both AR and ARX models predict test data more accurately and explain larger amounts of variability in the test data than a persistence model (Table 7). While AR and ARX models for BPR similarly explain more variability in the test data and achieve lower RMSEs, they produce larger MAPEs. For both HHV and BPR, the improvements relative to the baseline persistence models are small. Further, compared to the AR models, the performances of ARX models that consider digester grade are marginally better for as-fired HHV and BPR and slightly worse for as-fired viscosity. That the bivariate forecasts (ARX models) for HHV and BPR are more accurate than univariate forecasts (AR models) suggests that digester grade significantly influences HHV and BPR in this dataset (Tsay 2013).

For viscosity, the type of wood pulped likely impacts the viscosity of as-fired liquor, but a significant effect could not be seen in this dataset with our methodology. One reason may be due to viscosity having more “inertia” compared to HHV and BPR. As will be shown later, inertial units such as liquor storage tanks can introduce dynamics that diminish

the observable effect of upstream disturbances on downstream measurements. The dynamics of the effect of species on each as-fired liquor property probably differs. For example, below 50 % solids content, removing lignin from black liquor affects viscosity but not BPR (Moosavifar et al. 2006). As a result, the observable effect of species change on viscosity may be more diminished than those of HHV or BPR due to different process dynamics. This could also be exacerbated by the mill’s control system, whose goal is to suppress observable variations. Open-loop testing may provide the necessary data to observe the effect of species change on viscosity. However, it is likely impractical to perform open-loop testing on such a large area of the mill, which includes large, important units such as the digester and evaporators.

Finally, many factors, in addition to species, have been shown to affect black liquor viscosity such as kappa number and ash content (Zaman and Fricke 1996). It is possible that the impacts of other process operations on viscosity such as digester operation and ash recycling compete to be seen against the impact of species. Perhaps, with additional exogenous inputs, an ARX model would be able to identify the impact of species change on viscosity while accounting for the impacts of other potential exogenous inputs.

4.4 Similarity between ARX and AR model performances

Despite marginal improvements in prediction, the ARX models provide process insight: the lag between a change in species of wood pulped and the resulting change in as-fired liquor properties can be estimated. This demonstrates

Table 6: Prediction test data metrics on two tank simulation data (1 min predictions).

	Feed – tank 1 outlet			Feed – tank 2 outlet			Tank 1 outlet – tank 2 outlet		
	R^2	MAPE (%)	RMSE	R^2	MAPE (%)	RMSE	R^2	MAPE (%)	RMSE
Persistence	>0.99	0.17	1.05×10^{-3}	>0.99	0.13	7.97×10^{-4}	>0.99	0.13	7.97×10^{-4}
AR	>0.99	0.033	2.04×10^{-4}	>0.99	0.0049	3.59×10^{-5}	>0.99	0.0049	3.59×10^{-5}
ARX	>0.99	0.012	9.00×10^{-5}	>0.99	0.0046	3.45×10^{-5}	>0.99	0.0044	3.36×10^{-5}

Table 7: Prediction test data metrics on mill data (1 h predictions).

	HHV			Viscosity			BPR		
	R^2	MAPE (%)	RMSE	R^2	MAPE (%)	RMSE	R^2	MAPE (%)	RMSE
Persistence	0.70768	1.109	901.60	0.96618	6.092	5.867	0.86661	1.085	0.361
AR	0.72531	1.071	873.98	0.96619	6.081	5.866	0.86837	1.088	0.359
ARX	0.72749	1.066	870.52	0.96544	6.124	5.930	0.86994	1.089	0.357

how process prediction and process insight are distinct goals. If a model is only needed for prediction, the trade-off of increased model complexity for marginal improvements in prediction accuracy by the ARX model may be unnecessary. However, if the goal is to understand the relationship between an input and an output, then the ARX model is preferred: the AR model does not describe the relationship between an input and output.

The similar performance between ARX and AR models is likely due to the heavy autocorrelation of process data. As mentioned earlier, ARX models can be thought of as a combination of a filtered input and filtered white noise. However, if the output Y_t is heavily autocorrelated (i.e., the coefficients a_i in Equation (9) are large), X_t will do little to boost prediction accuracy. Both the simulation data and the mill data in this work, like most process data, are highly autocorrelated.

To demonstrate the affect of autocorrelation, consider the two-tank simulation (Figure 1). If C_0 is the concentration of NaCl in the feed stream; C_1 is the NaCl concentration in the outlet of tank 1; C_2 is the concentration of NaCl in the outlet of tank 2; Q is the volumetric flowrate; V_1 is the volume of fluid tank 1; V_2 is the volume of fluid in tank 2; and perfect mixing is assumed, then the dynamics of Tanks 1 and 2 are described by the equations

$$V_1 \frac{dC_1}{dt} = QC_0(t) - QC_1(t)$$

and

$$V_2 \frac{dC_2}{dt} = QC_1(t) - QC_2(t).$$

Now, discretize both equations and substitute $\tau_i = \frac{V_i}{Q}$:

$$\frac{\tau_1}{\Delta t} (C_{1,t} - C_{1,t-\Delta t}) = C_{0,t} - C_{1,t},$$

$$\frac{\tau_2}{\Delta t} (C_{2,t} - C_{2,t-\Delta t}) = C_{1,t} - C_{2,t},$$

which can be rearranged to obtain

$$C_{1,t} = \left(1 - \frac{1}{1 + \frac{\Delta t}{\tau_1}}\right) C_{0,t} + \frac{1}{1 + \frac{\Delta t}{\tau_1}} C_{1,t-\Delta t}$$

and

$$C_{2,t} = \left(1 - \frac{1}{1 + \frac{\Delta t}{\tau_2}}\right) C_{1,t} + \frac{1}{1 + \frac{\Delta t}{\tau_2}} C_{2,t-\Delta t},$$

respectively. Finally, substitute $\rho_i = \frac{1}{1 + \frac{\Delta t}{\tau_i}}$:

$$\begin{aligned} C_{1,t} &= (1 - \rho_1)C_{0,t} + \rho_1 C_{1,t-\Delta t} \\ C_{2,t} &= (1 - \rho_2)C_{1,t} + \rho_2 C_{2,t-\Delta t} \end{aligned} \quad (13)$$

The autocorrelations of $C_{1,t}$ and $C_{2,t}$ are ρ_1 and ρ_2 , respectively. Each autocorrelation depends on the hydraulic residence time τ_i and the sampling interval Δt . Large residence times relative to short sampling intervals ($\tau \gg \Delta t$) are typical in process data. As a result, mill measurements tend to be highly autocorrelated ($\rho_i \approx 1$). In the two tank system (Equation (13)), this causes the correlations between $C_{2,t}$ and upstream measurements, $C_{0,t}$ and $C_{1,t}$, to be small.

This suggests that, in practice, predictive models will benefit little from measurements upstream of units characterized by large residence times relative to the sampling interval. This does not imply that upstream properties are irrelevant: in the two-tank system, C_0 determines C_1 and C_2 . Rather, the observed correlation between downstream measurements and upstream disturbances depends on the magnitude of residence times as compared to the sampling interval.

As a result, the sampling interval is an important consideration when fitting models and identifying lags. If the interval is too large, the effects of autocorrelation would diminish, but relevant signal may go unseen. In contrast, if the interval is too small, the relevant signal may be buried in high-frequency noise. In this study, we used a sampling interval of 1 h, which aligns with suggestions by Box et al. (2015) and Ljung (1998) that sampling intervals for chemical processes should be on the order of hours. When a shorter sampling interval (30 min) or a longer one (4 h) was used, the identified ARX models for HHV and BPR produced unrealistic lags and failed to outperform the AR models. For viscosity, as with the 1 h sampling interval, ARX models underperformed the AR model. This highlights the importance of comparing the bivariate models and the univariate models on unseen test data as a final check. Additionally, a shorter sampling interval should be favoured. Since lags are estimated as a number of timesteps, a shorter sampling interval provides a more precise lag estimate.

5 Conclusions

Simple ARX models seem to capture process lags in the kraft recovery process. Here, we validated the use of ARX models to identify lags in a simulation of two tanks in series. Then, we fit ARX models to data from a North American mill, estimating the lagged effect of a change in species pulped on three as-fired liquor properties: HHV, viscosity, and BPR. The estimated lags for BPR (41 h) and as-fired HHV (49 h) agree with a detailed simulation of the mill and are close to the estimated hydraulic residence times through the same system. The effect of a change in digester grade on as-fired

viscosity could not be identified and is suspected to be masked by noise. While the ARX models provide process insight, they provide marginal improvements over AR models and the persistence model for process prediction. This is likely due to the autocorrelation of mill measurements. We derive this result for the case of species concentration throughout two tanks in series. This suggests that, in practice, measurements upstream of units characterized by large residence times provide little benefit to predictive models.

Research ethics: Not applicable.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission. Jerry Ng: methodology, data acquisition, analysis, writing. Yuri Lawryshyn: supervision, editing. Nikolai DeMartini: supervision, editing, funding.

Competing interests: The authors state no conflict of interest.

Research funding: 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.

Data availability: The raw data can be obtained on request from the corresponding author.

References

- Adams, T.N. (1997). *Kraft recovery boilers*. TAPPI Press, Atlanta, Georgia, United States.
- Almeida, G.M.D. and Park, S.W. (2013). Time-delay neural network for monitoring of abnormal situations in continuous industrial processes. Case study: evaporation station of a pulp mill. In: *IFAC proceedings volumes, volume 46 of 11th IFAC workshop on intelligent manufacturing systems*, pp. 408–413, Available at: <https://www.sciencedirect.com/science/article/pii/S1474667015357098>.
- Aurel Systems (2019), *CADSIM Plus*, Available at: www.aurelsystems.com/cadsim-plus/
- Bisgaard, S. and Kulahci, M. (2011). *Time series analysis and forecasting by example*. John Wiley & Sons, Hoboken, New Jersey, United States.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M. (2015). *Time series analysis: forecasting and control*, 5th ed. John Wiley & Sons, Hoboken, New Jersey, United States.
- Chen, X. and Zhao, C. (2020). Multivariate time delay estimation based on dynamic characteristic analytics. In: *2020 39th Chinese control conference (CCC)*, pp. 2306–2311.
- Chetouani, Y. (2008). *Using ARX approach for modelling and prediction of the dynamics of a reactor-exchanger*, 154. Institution of Chemical Engineers, Rugby, United Kingdom, p. 297.
- Correia, F.M., d'Angelo, J.V.H., Almeida, G.M., Mingoti, S.A., Correia, F.M., d'Angelo, J.V.H., Almeida, G.M., and Mingoti, S.A. (2018). Predicting kappa number in a kraft pulp continuous digester: a comparison of forecasting methods. *Braz. J. Chem. Eng.* 35: 1081–1094, <https://doi.org/10.1590/0104-6632.20180353s20160678>.
- Dahlbäck, M. (2020). Evaluation of models for process time delay estimation in a pulp bleaching plant, Master's thesis, Umea University.
- Ekbåge, D., Nilsson, L., and Håkansson, H. (2019). Time series analysis of refining conditions and estimated pulp properties in a chemi-thermomechanical pulp process. *Bioresources* 14: 5451–5466, <https://doi.org/10.15376/biores.14.3.5451-5466>.
- El-Din, A.G. and Smith, D.W. (2002). A combined transfer-function noise model to predict the dynamic behavior of a full-scale primary sedimentation tank. *Water Res.* 36: 3747–3764, [https://doi.org/10.1016/S0043-1354\(02\)00089-1](https://doi.org/10.1016/S0043-1354(02)00089-1).
- Ettaleb, L., Dumont, G.A., and Davies, M.S. (1998). An extended off-line least-squares method for parameter identification and time delay estimation. In: *Proceedings of the 37th IEEE conference on decision and control (Cat. No.98CH36171)*, Vol. 3, pp. 3423–3428.
- Galicia, H.J., Peter He, Q., and Wang, J. (2011). A reduced order soft sensor approach and its application to a continuous digester. *J. Process Control* 21: 489–500, <https://doi.org/10.1016/j.jprocont.2011.02.001>.
- Graziani, S. and Xibilia, M.G. (2019). Design of a soft sensor for an industrial plant with unknown delay by using deep learning. In: *2019 IEEE international instrumentation and measurement technology conference (I2MTC)*, pp. 1–6.
- Graziani, S. and Xibilia, M.G. (2021). On the use of correlation analysis in the estimation of finite-time delay in soft sensors design. In: *2021 IEEE international instrumentation and measurement technology conference (I2MTC)*, pp. 1–6.
- Hart, P.W. (2022). *Chemical recovery in the alkaline pulping processes*, 4th ed. TAPPI, Peachtree Corners, Georgia, United States.
- He, X. and Asada, H. (1993). A new method for identifying orders of input-output models for nonlinear dynamic systems. In: *1993 American control conference*, pp. 2520–2523.
- John, M. and Ferbinteanu, J. (2021). Detecting time lag between a pair of time series using visibility graph algorithm. *Commun. Stat., Case Stud. Data Anal. Appl.* 7: 315–343, <https://doi.org/10.1080/23737484.2021.1882356>.
- Lee, G., Tosukhowong, T., Lee, J.H., and Han, C. (2006). Fault diagnosis using the hybrid method of signed digraph and partial least squares with time delay: the pulp mill process. *Ind. Eng. Chem. Res.* 45: 9061–9074, <https://doi.org/10.1021/ie060793j>.
- Ljung, L. (1998). *System identification*, 2nd ed. Prentice Hall PTR, Upper Saddle River, New Jersey, United States.
- Matos, A.S., Requeijo, J.G., and Pereira, Z.L. (2008). Integration of engineering process control and statistical control in pulp and paper industry. In: Braunschweig, B. and Joulia, X. (Eds.). *Computer aided chemical engineering, volume 25 of 18 European symposium on computer aided process engineering*. Elsevier, pp. 399–404, Available at: <https://www.sciencedirect.com/science/article/pii/S1570794608800715>.
- Mechaqrane, A. and Zouak, M. (2004). A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building. *Neural Comput. Appl.* 13: 32–37, <https://doi.org/10.1007/s00521-004-0401-8>.
- Miikkulainen, P., Kankkunen, A., Järvinen, M.P., and Fogelholm, C.J. (2005). Predicting droplet size from black liquor spray characteristics. *TAPPI J.* 4(5): 11–17.
- Moosavifar, A., Sedin, P., and Theliander, H. (2006). Viscosity and boiling point elevation of black liquor: consequences when lignin is extracted from the black liquor. *Nord. Pulp Pap. Res. J.* 21: 180–187, <https://doi.org/10.3183/npprj-2006-21-02-p180-187>.

- Rebello, C.M., Marrocos, P.H., Costa, E.A., Santana, V.V., Rodrigues, A.E., Ribeiro, A.M., and Nogueira, I.B.R. (2022). Machine learning-based dynamic modeling for process engineering applications: a guideline for simulation and prediction from perceptron to deep learning. *Processes* 10: 250, <https://doi.org/10.3390/pr10020250>.
- Rogerson, A.M. (2022). Dynamic modelling of process chemistry in kraft pulp mills, Master's thesis, Available at: <https://tspace.library.utoronto.ca/handle/1807/124895> (Accessed 11 November 2022).
- Sainlez, M. and Heyen, G. (2013). Comparison of supervised learning techniques for atmospheric pollutant monitoring in a kraft pulp mill. *J. Comput. Appl. Math.* 246: 329–334, <https://doi.org/10.1016/j.cam.2012.06.026>.
- Sayda, A.F. and Taylor, J.H. (2003). Model predictive control of a mechanical pulp bleaching process. *IFAC Proc. Vol.* 36: 293–298, [https://doi.org/10.1016/S1474-6670\(17\)33341-4](https://doi.org/10.1016/S1474-6670(17)33341-4).
- Souza, F. and Araújo, R. (2011). Variable and time-lag selection using empirical data. In: *ETFA2011*, pp. 1–8.
- The MathWorks Inc (2022). *System identification toolbox*, Available at: <https://www.mathworks.com>.
- Topalian, S.O.N., Ramin, P., Kjellberg, K., Kulahci, M., Flores Alsina, X., Batstone, D.J., and Gernaey, K.V. (2022). Forecasting operational conditions: a case-study from dewatering of biomass at an industrial wastewater treatment plant. In: Yamashita, Y. and Kano, M. (Eds.). *Computer aided chemical engineering, volume 49 of 14 international symposium on process systems engineering*. Elsevier, pp. 2077–2082, Available at: <https://www.sciencedirect.com/science/article/pii/B9780323851596503468>.
- Tsay, R.S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons, Hoboken, New Jersey, United States.
- Wang, J.S. and Chen, Y.P. (2006). A fully automated recurrent neural network for unknown dynamic system identification and control. *IEEE Trans. Circuits Syst. I: Regul. Pap.* 53: 1363–1372, <https://doi.org/10.1109/TCSI.2006.875186>.
- Wasson, R. (2016). Development of a supervisory control for a kraft pulp mill steam and power system using PCA and PLS, Master's thesis, University of New Brunswick, Available at: <https://unbscholar.lib.unb.ca/islandora/object/unbscholar>.
- Zaman, A.A. and Fricke, A.L. (1996). Effect of pulping conditions and black liquor composition on Newtonian viscosity of high solids kraft black liquors. *Ind. Eng. Chem. Res.* 35: 590–597, <https://doi.org/10.1021/ie950202h>.
- Zenger, K. and Ylinen, R. (1994). Simulation of variable delays in material transport models. *Math. Comput. Simulat.* 37: 57–72, [https://doi.org/10.1016/0378-4754\(94\)90059-0](https://doi.org/10.1016/0378-4754(94)90059-0).