Online Monitoring of Inner Deposits in Crude Oil Pipelines

Riccardo Giro[1]*, Giancarlo Bernasconi[1], Giuseppe Giunta[2], and Simone Cesari[2]

[1]Politecnico di Milano
[2]Eni S.p.A

Summary

The formation of deposits is a very common issue in oil and gas pipeline transportation systems. Such sediments, mainly wax and paraffine for crude oil, or hydrates and water for gas, progressively reduce the free cross-sectional area of the pipe, leading in some cases to the complete occlusion of the conduit. The overall result is a decrease in the transportation performance, with negative economic, environmental, and safety consequences. To prevent this issue, the amount of inner deposits must be continuously and accurately monitored, such that the corresponding cleaning procedures can be performed when necessary. Currently, the former operation is still dictated by best-practice rules pertaining to preventive or reactive approaches, yet the demand from the industry is for predictive solutions that can be deployed online for real-time monitoring applications. The paper moves toward this direction by presenting a machine learning methodology that leverages pressure measurements to perform online monitoring of the inner deposits in crude oil trunklines. The key point is that the attenuation of pressure transients within the fluid is dependent on the free cross-sectional area of the pipe. Pressure signals, collected from two or more distinct locations along a pipeline, can therefore be exploited to estimate and track in real time the presence and thickness of the deposits. Several statistical indicators, derived from the attenuation of such pressure transients between adjacent acquisition points, are fed to a data-driven regression algorithm that automatically outputs a numeric indicator representing the amount of inner pipe debris. The procedure is applied to the pressure measurements collected for one and a half years on discrete points at a relative distance of 40 and 60 km along an oil pipeline in Italy (100 km length, 16-in. inner diameter pipes). The availability of historical data prepipe and postpipe cleaning campaigns further enriches the proposed data-driven approach. Experimental results demonstrate that the proposed predictive monitoring strategy is capable of tracking the conditions of the entire conduit and of individual pipeline sections, thus determining which portion of the line is subject to the highest occlusion levels. In addition, our methodology allows for real-time acquisition and processing of data, thus enabling the opportunity for online monitoring. Prediction accuracy is assessed by evaluating the typical metrics used in the statistical analysis of regression problems.

Introduction

Pipeline transportation systems represent the cheapest and safest solution to convey hydrocarbons, gases, and other fluids over long distances. After construction, pipe internals tend to naturally accumulate deposits, such as rust, dirt, mill scale, or paraffin wax (McAllister 2013). Those constituents need to be removed for a number of reasons: First, to avoid product contamination, which can have a negative economic impact on the business; second, to allow for a better use of corrosion inhibitors, whose action is less effective if the pipe bore is covered with mill scale or it is partially corroded; third, to improve flow rate and efficiency, which is maximized when the pipeline is completely clean (especially for pipelines having a length of several tens of kilometers); and lastly, in natural gas transportation lines, to facilitate pipeline drying, required to prevent internal corrosion and the formation of hydrates.

The internal cleaning of a pipeline can be performed with several techniques, applied individually or jointly (McAllister 2013; Olajire 2021). We mention here the injection of chemical solvents (e.g., flux); internal sandblasting, in which an abrasive material is used to scrape the inner surface of the pipe and to remove contaminants; purging with air or gas to prevent oxidation phenomena leading to corrosion; running a pipeline inspection gauge (PIG), which is a multipurpose maintenance tool capable of flushing debris out of the pipe by scraping its internals with metallic brushes or plastic disks. Among those cleaning solutions, the PIG becomes particularly advantageous from a product saving and environmental point of view, especially in multiproduct lines where fluids are conveyed in batches: For instance, at the end of a given oil transfer, one can clear out the residuals stuck inside the pipe bore with a PIG run, thus allowing for a faster and effective product switch. In fact, by separating two batches of different products with a PIG, it is possible to avoid flushing the line with water, solvents, or (in some cases) the following product, and to stave off effluent treatment or contaminated product recovery.

Regardless of the solution used, adequate strategies should be carefully arranged to prevent blockages in pipes and to guarantee the desired transportation efficiency. For these reasons, tracking the amount of inner deposits assumes particular relevance in pipeline transportation systems (Van der Geest et al. 2021), especially for crude oil lines where the high viscosity of the conveyed product facilitates the deposition of wax and asphaltenes. To this date, however, this monitoring operation is still performed by resorting to empirical, best-practice rules or by scheduling a periodic activity, an approach used in preventive or reactive maintenance: Such a policy is mainly adopted because we lack both a clear definition of clean/clogged pipe and a rigorous method for measuring the related accumulation of debris. A significant body of literature focuses the attention on the prediction and monitoring of inner deposits in oil and gas pipelines. Several authors have proposed deterministic deposition models, yet they can only be applied in an offline context (Giacchetta et al. 2019; Leporini et al. 2019; Xie et al. 2018; Kamari et al. 2013, 2014; Shasha and Qiyu 2014; Wang et al. 2014b; Obaseki and Elijah 2021; Yao et al. 2021; Chen et al. 2021); multiple studies have instead been designed and tested on laboratory setups or make use of data from the literature, and so they lack a validation phase on real scenarios (Huang and Ma 2008; Huang et al. 2017; Guozhong and Gang 2010; Ito et al. 2021; Li et al. 2020; Modesty Kelechukwu et al. 2013; Wang and Huang 2014; Wang et al. 2014a; Zougari 2010; Li et al. 2018; Theyab and Diaz 2016; Van Der Geest et al. 2018; Adeyanju and Oyekunle 2019; Chi et al. 2019; Sun et al. 2020; Obinichi et al. 2021;

Li et al. 2021; Agarwal et al. 2021). According to the authors' knowledge, there are currently two research works that satisfy the requirement of real-time monitoring: The former, by Halstensen et al. (2013), is an online estimation method based on acoustic chemometrics, which has, however, been validated on a very short pipe section (5.5 m of length); the latter, by Lock Sow Mei et al. (2015), is a technique based on electrical capacitance tomography, yet it has been designed and tested in a laboratory. Lastly, certain authors have implemented data-driven solutions based on machine learning approaches (Xie and Xing 2017; Obanijesu and Omidiora 2008; Jalalnezhad and Kamali 2015; Sousa et al. 2021; Menad et al. 2021), but they each share one or more of the previously discussed drawbacks.

The literature review reported here highlights two main research gaps: First, there is a need for consistent and precise prediction methods to monitor the amount of inner deposits in pipelines, making use of flexible data-driven approaches to be validated on real data sets (Mwendapole Lonje and Liu 2021); second, modern systems demand online monitoring to perform real-time predictive maintenance (Alnaimat and Ziauddin 2020), yet this requirement is typically not satisfied by the current research. This work addresses all the aforementioned necessities by presenting a machine learning methodology (based on extremely randomized trees) that makes use of pressure measurements, collected in two or more discrete points along a pipeline, to automatically provide as output a numeric indicator that quantifies the cleanliness level of the pipeline itself, thus offering a clear indication of its internal conditions. We demonstrate that our proposal can track the occlusion levels of an entire line and of individual pipe segments, therefore determining which portion of the conduit is mostly blocked by deposits and debris. In addition, the methodology presented here can operate with data collected in real time from transportation assets, being thus capable of performing online monitoring and control tasks. Lastly, the validity of the proposed procedure has been assessed on one and a half years of data, collected from a 100km crude oil pipeline located in Italy.

The remainder of the paper is structured as follows. We first outline the proposed prediction method. Then, we describe in detail its application on a real crude oil pipeline. Lastly, we draw the conclusions.

Overview of the Prediction Method

The approach presented here for monitoring deposits in crude oil transportation pipelines makes use of standard pressure measurements, collected by means of hydrophones in two distinct points (A, B) along a pipeline; such instruments sense the pressure transients propagating within the fluid that is flowing inside the pipe. Acoustic signals can be generated by multiple sources, such as pumping equipment, valves, flow turbulence, any PIG traveling inside the line, spill operations, tremors, quakes, landslides, etc.; in this particular case, the main emitters of interest are represented by the pumps and by the PIGs (Bernasconi and Giunta 2020).

The raw pressure measurements are suitably processed to compute the specific attenuation $\alpha_{\overline{AB}}$ of acoustic waves propagating between A and B. Successively, a set of statistical indicators x is evaluated from the aforementioned quantity, and the resulting feature set is used to train a machine learning algorithm based on extremely randomized trees (Geurts et al. 2006). The latter is designed to output a real number ranged between zero and unity, which describes the current state of pipeline internals (clean = 0, dirty = 1 or any intermediate stage). As a last step, it follows an assessment phase, in which the accuracy of the data-driven prediction model is tested on unseen data. If the performance of the predictor satisfies the design requirements (e.g., accuracy metrics greater than a target threshold), the model can be deployed online to monitor in real time the inner conditions of the pipeline. In such a case, at each timestep k, one has to perform the following operations:

1. Collect the instantaneous pressure measurements $P_{A,k}$ and $P_{B,k}$ from the locations A and B, respectively;
2. Compute $\alpha_{\overline{AB},k}$
3. Evaluate the required set of features, denoted by the vector x_k
4. Obtain the instantaneous prediction $\hat{y}_k$ by providing x_k as input to the regression algorithm.

Application of the Model on a Real Transportation Scenario: Chivasso-Pollein Crude Oil Pipeline

This section demonstrates how the proposed model has been applied to data collected from a real pipeline transportation system. Two applications will be presented: The first one consists of a global monitoring strategy to predict the occlusion levels of a conduit in its entirety, whereas the latter is used to monitor the amount of deposits within individual line sections.

Experiment Setup. We have used the historical vibroacoustic measurements, collected by a proprietary digital integrity monitoring system (e-vpms® technology; Giunta and Bernasconi 2019; Giunta et al. 2016), installed on a crude oil transportation line that connects the Eni logistic terminals of Chivasso and Pollein, located in north Italy (Bernasconi et al. 2014). Such a line has a length of approximately 100 km and is characterized by 16-in. inner diameter pipes. A schematic representation of the e-vpms® system is displayed in **Fig. 1.** A set of sensing stations are located in discrete points (named A, B, and C in **Fig. 1**) along the pipeline; each e-vpms® acquisition unit is equipped with a sensing group, recording the absolute pressure of the transported fluid in bars, and a dynamic hydrophone, which measures small-scale dynamic pressure variations (in the order of kilopascals). The collected measurements are time synchronized by means of GPS and, successively, sent to a central control unit. Pressure data have been collected at a sampling rate of 20 Hz from 1 June 2013 until 1 December 2014. In addition to the e-vpms® data set, we also dispose of historical prepipe and postpipe cleaning operation logs, which textually outline the dates and times at which one or more PIG runs have been performed on the trunkline.

The satellite map of the conduit and the location of the recording stations (labeled with the letters A, B, and C) are displayed in **Fig. 2,** respectively, with a red line and yellow pins. The distances between each station and the pumping equipment located at Terminal A are reported in **Table 1** (Giro et al. 2021).

Station	Distance with Respect to Station A (km)
A	0
B	59.3
C	100.4

Table 1—Distance between each e-vpms® station and the pumping terminal positioned in Station A.
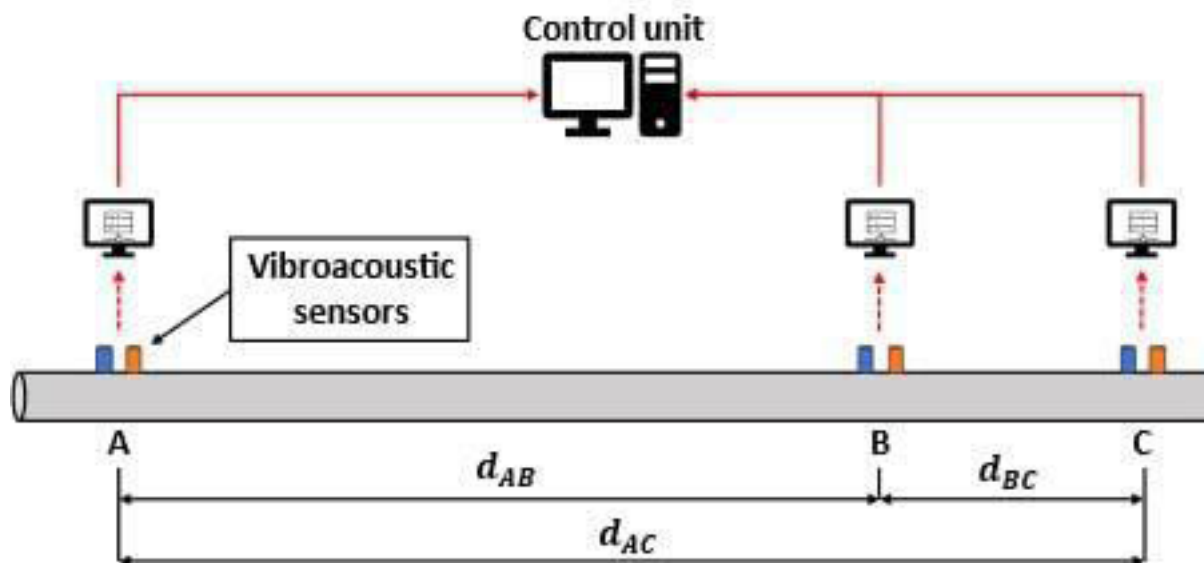
Fig. 1—Scheme of the e-vpms® vibroacoustic monitoring system.

Data Processing. The first step consists in transforming the unprocessed measurements into a format suitable for machine learning tasks. The two plots in **Fig. 3,** respectively, show the raw static (**Fig. 3**, top) and dynamic (**Fig. 3**, bottom) pressure signals, collected from the three different e-vpms® stations (A, B, and C, respectively identified with turquoise, purple, and green lines). Each pressure time series needs to be cleansed to remove undesired data points:

1. Presence of outliers because of sensor errors. Such outliers are because of rare electromagnetic disturbances affecting the power unit of the measuring stations and result in faulty acquisitions having values outside of the dynamic range of the instrumentation. In this specific case, static pressure readings lower than 0.5 bar and higher than 80 bar are discarded; likewise, dynamic pressure values below −170 kPa and above 170 kPa are eliminated from the data set.
2. Unwanted pressure values corresponding to operational statuses of the line not contributing significantly to the occlusion levels of the pipes. More specifically, we assume that the formation of deposits within pipe segments mainly occurs when the oil is actively conveyed through the line. In fact, at the end of each batch, the pipeline is filled with a flux. Therefore, all the corresponding time intervals in which the pipeline is not operational (e.g., off) or it is into a flow regulation state (e.g., pressure transients generated by pumping fluctuations) should be ruled out from the data set.

Even though each of these impairments can be addressed manually, tackling Step 2 by hand becomes impractical when processing data sets having billions of points. A possible solution to this problem consists in exploiting some automated detection procedure, such as the data-driven pump monitoring system described in Giro et al. (2021); Giunta et al. (2020). We have therefore applied the clustering method outlined in (Giro et al. 2021) to fit a Gaussian mixture model (GMM) to the available pressure data. The Gaussian mixture model automatically produces as output a set of categorical labels, each indicating all the time instants in which the system is either off or is performing flow regulations; therefore, the corresponding data points can be easily identified and removed. It should be stressed that the static pressure data have been only used here to aid and simplify the completion of Step 2; however, having such measurements at disposal is not mandatory at all for the purpose of monitoring inner deposits, and from this point onward, the discussion will be solely focused on the analysis of the pressure transients (e.g., dynamic pressure data). Lastly, **Fig. 4** represents the static (**Fig. 4**, top) and dynamic (**Fig. 4**, bottom) pressure measurements after having performed the processing steps previously described.
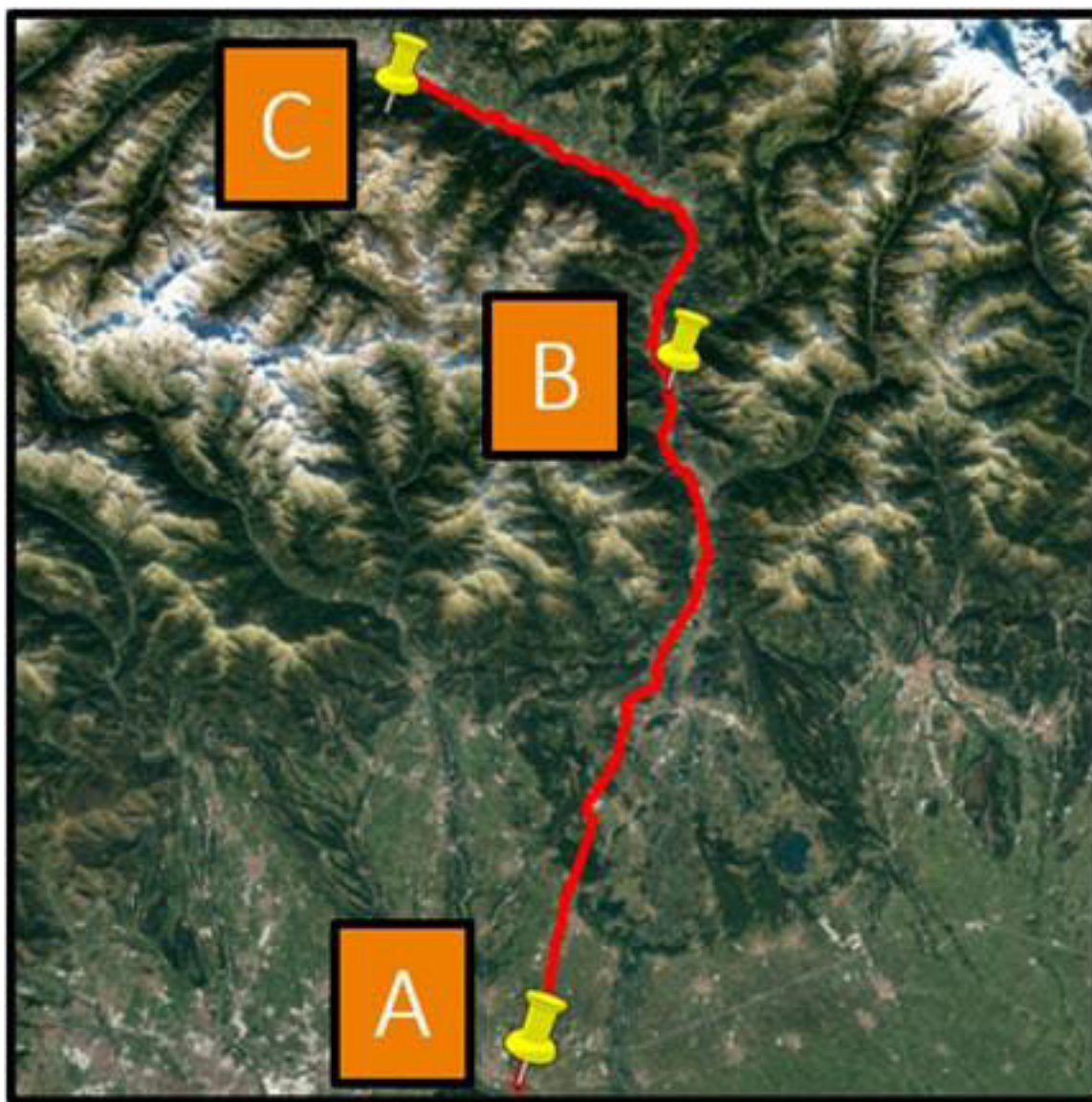
Attenuation Analysis. We demonstrate here that attenuation measurements (validated by PIG tracking) prove to be a valuable feature for assessing the inner status of the pipeline. For single-phase fluids, the specific attenuation α of acoustic waves propagating within the pipe can be expressed as follows (Blackstock and Atchley 2001):

$$\alpha = \frac{1}{a}\sqrt{\frac{\pi f \mu}{\rho v^2}}\ \left[\frac{dB}{km}\right], \tag{1}$$

where a is the internal radius of the pipe (m), f corresponds to the frequency (Hz), μ is the dynamic viscosity of the fluid (Pa·s), ρ refers to the fluid density (kg/m^3), and v is the measured sound speed within the fluid (m/s). Of all these parameters, particular attention should be given to a, as it experiences the most significant variations during cleanup campaigns. Before such operations, pipe sections are internally affected by deposits (especially wax), thus reducing the effective internal diameter a of the pipe in which the oil can flow. It follows that the specific attenuation of acoustic waves in a pipe segment is greater when the latter is partially clogged by wax, compared with a clean pipe (e.g., after cleaning operations).

Instead of using (1) to measure α, which requires the real-time knowledge of the instantaneous parameters of the fluid (e.g., μ, ρ, and v), we have developed a novel and simpler approach to derive the specific attenuation of acoustic waves, which only makes use of basic pressure transients. The grid plot of **Fig. 5** graphically explains the aforementioned statement with an example: **Fig. 5** displays the attenuation levels a few days before and after a PIG campaign performed on 13 May 2014, as recorded in the available maintenance logs. Starting from the time series of dynamic pressure (**Fig. 5**, charts on the first column), one can evaluate the power spectral density of such signals collected at two different stations (in this example, A and C: **Fig. 5**, plots on the second column) and successively derive the frequency-dependent specific attenuation values as the ratio between the two power spectral densities, divided by the distance (in km) between the two stations (**Fig. 5**, charts on the third column). Lastly, the average specific attenuation level can be obtained by integrating
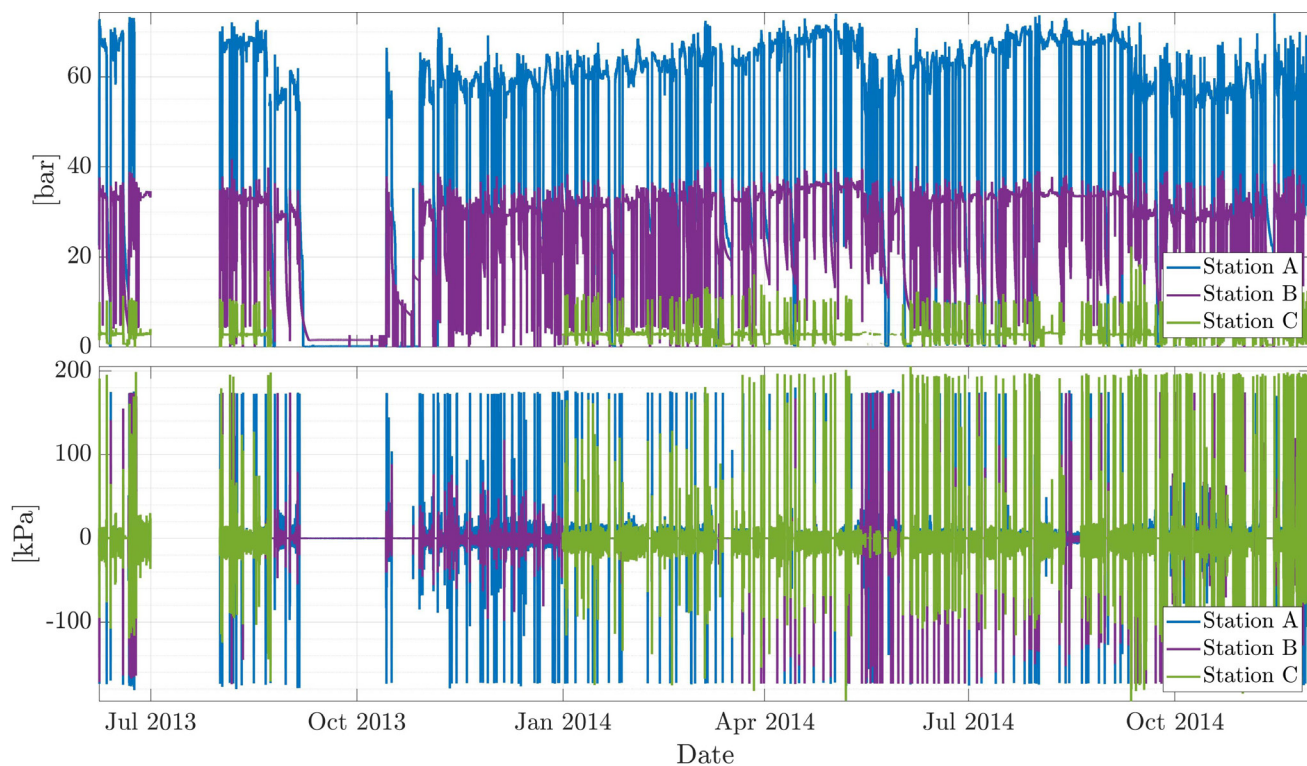
Fig. 2—Satellite map of Chivasso-Pollein pipeline routing (red curve) and location of the e-vpms® measurement stations (yellow pins).
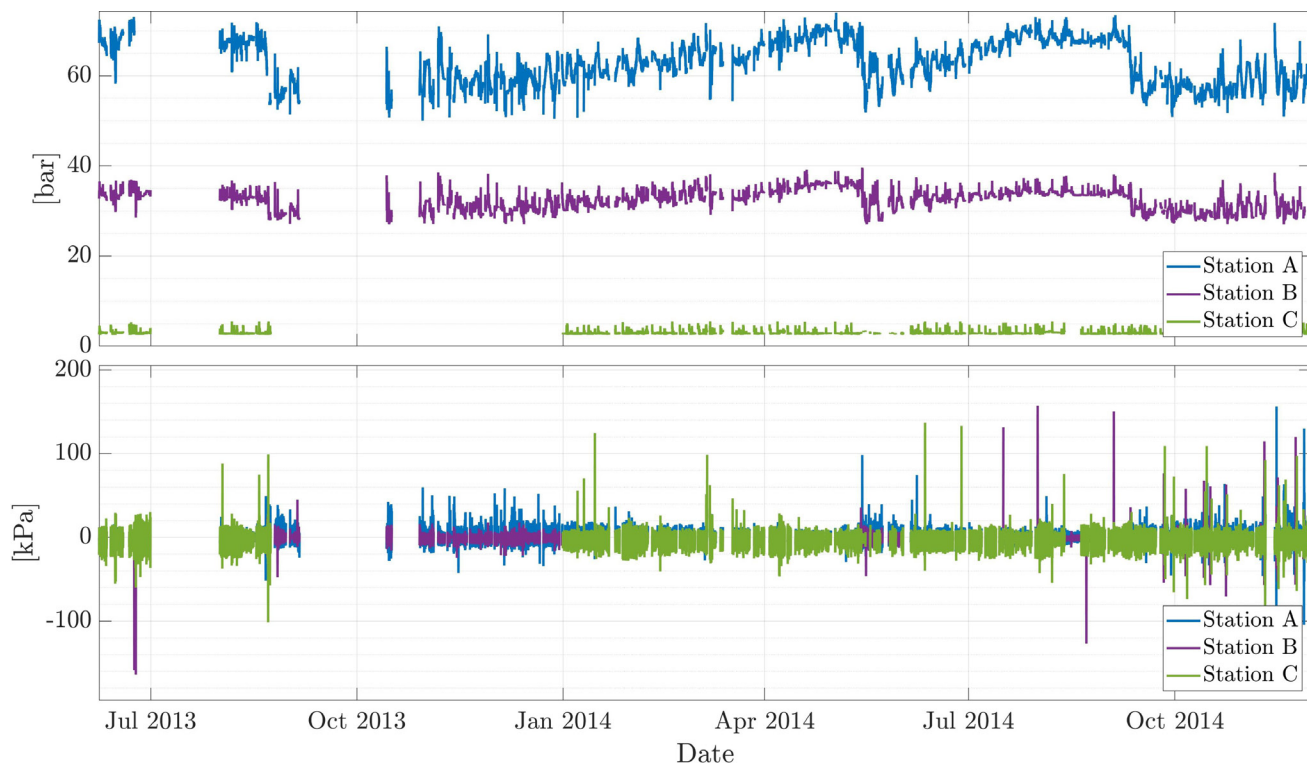
α within the entire frequency range (in our case, between 0 and 10 Hz); in other words, α corresponds to a power ratio between two signals, scaled by a distance factor.

The procedure described above can be periodically executed to derive the temporal evolution of the specific attenuation for any line section. An example is depicted in **Fig. 6,** where the short-term attenuation value $\alpha_{\overline{AC}}$ (gray line) is displayed for the $\overline{AC}$ line segment. The long-term trend (magenta curve) has instead been derived from the short-term values by smoothing the latter curve with a noncausal, 1 week moving average. We can observe that the long-term attenuation curve is characterized by a slow and gradual increase over the course of several weeks or months, coupled with rapid decreases having a much shorter temporal duration: The former phenomenon is mainly because of a progressive augmentation in the occlusion levels of the pipes; the latter correspond to the pigging campaigns performed on the pipeline (as reported in the maintenance logs), some of which have been highlighted in **Fig. 6** using black vertical bars. It can be noted that every major drop in attenuation occurs right after a pigging operation has been executed. In such circumstances, cleaner pipe sections allow the pumping terminal located at Station A to operate with a lower service pressure while still delivering the reference value of about 3 bar at Station C (as it can be inferred from the topmost plot of **Fig. 4**).

Validation through PIG Detection and Tracking. We have observed that the attenuation represents a valuable indicator of a pipe's occlusion level. The goodness of such a feature can be further validated if one experimentally verifies the occurrence of each PIG campaign (as reported in the available operation logs), especially for the main operations (highlighted in **Fig. 6** using black vertical bars). To do so, we have developed a software tool capable of detecting, in the observed dynamic pressure measurements, the acoustic noise generated by the traveling PIG (Bernasconi and Giunta 2020). An example of the output provided by such a software is displayed in **Fig. 7,** where the positions of several PIGs inside the pipeline have been tracked during the second half of November 2014. If we consider the topmost chart of **Fig. 7**, the latter represents a density plot of the normalized cross-correlation $R_{\overline{AC}}$ between the dynamic pressure transients recorded by the hydrophones at Stations A and C, as a function of time (horizontal axis) and of the distance $d_A(x)$ from Station A (vertical axis). Darker regions of the image correspond to values of $R_{\overline{AC}}$ closer to unity (maximum correlation); similarly, lighter areas
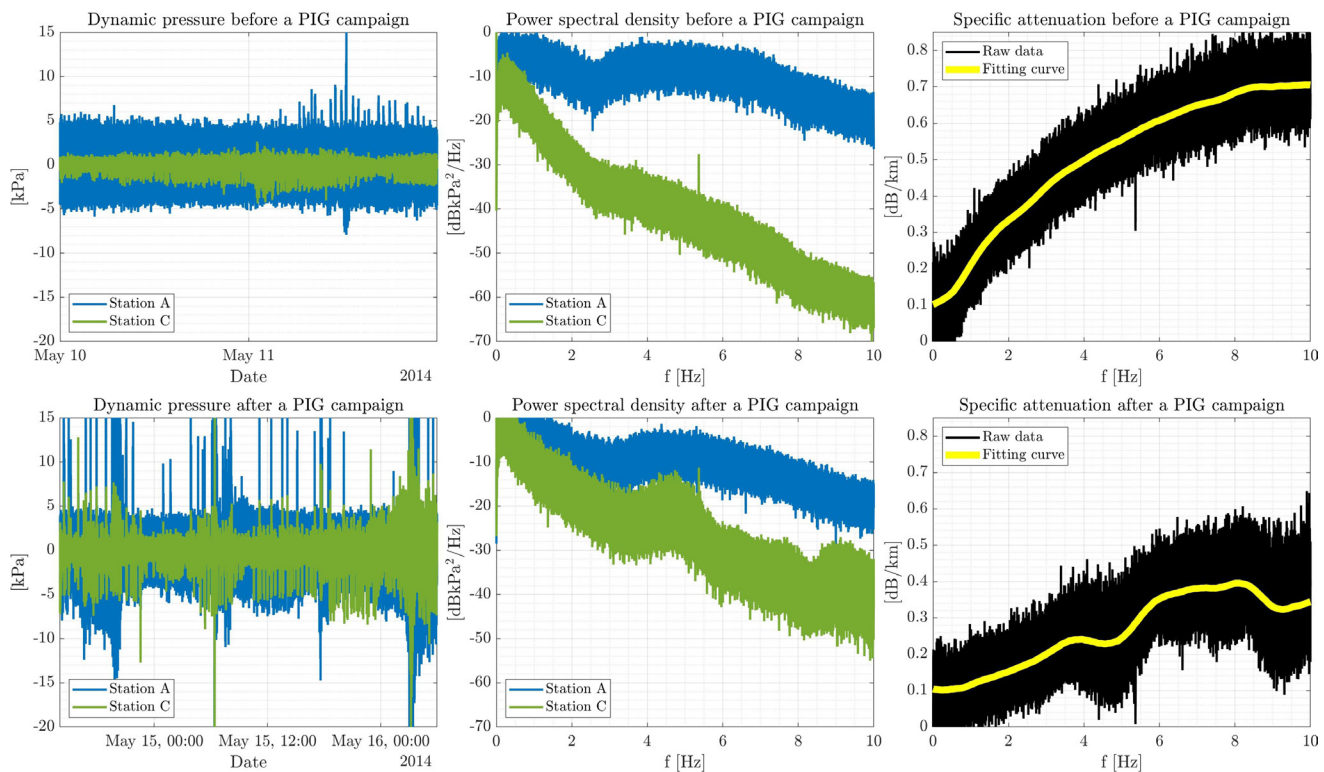
Fig. 3—Raw static (top) and dynamic (bottom) pressure time series, as measured from Stations A, B, and C (respectively, colored with turquoise, purple, and green lines).
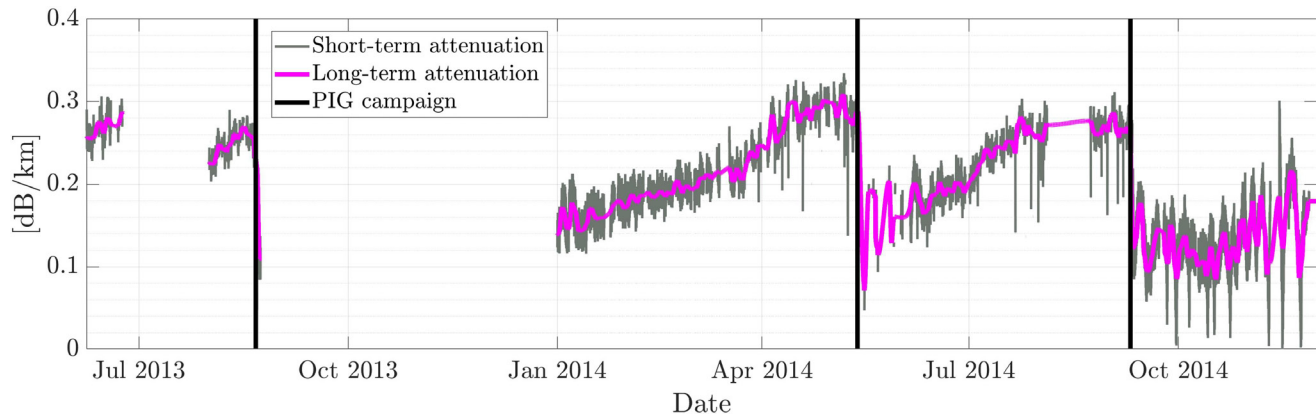

Fig. 4—Processed static (top) and dynamic (bottom) pressure time series, as measured from Stations A, B, and C (respectively, colored with turquoise, purple, and green lines).

present the lowest cross-correlation values, which tend to be zero. For simplicity, we have converted the correlation delays $\tau_A(x)$ (that would be displayed on the vertical axis) into distances, because the former are a linear function of the latter and of the sound velocity v within the fluid conveyed in the pipe:

Fig. 5—Dynamic pressure signals time series, power spectral density, and specific attenuation curves before (top row) and after (bottom row) a PIG campaign.
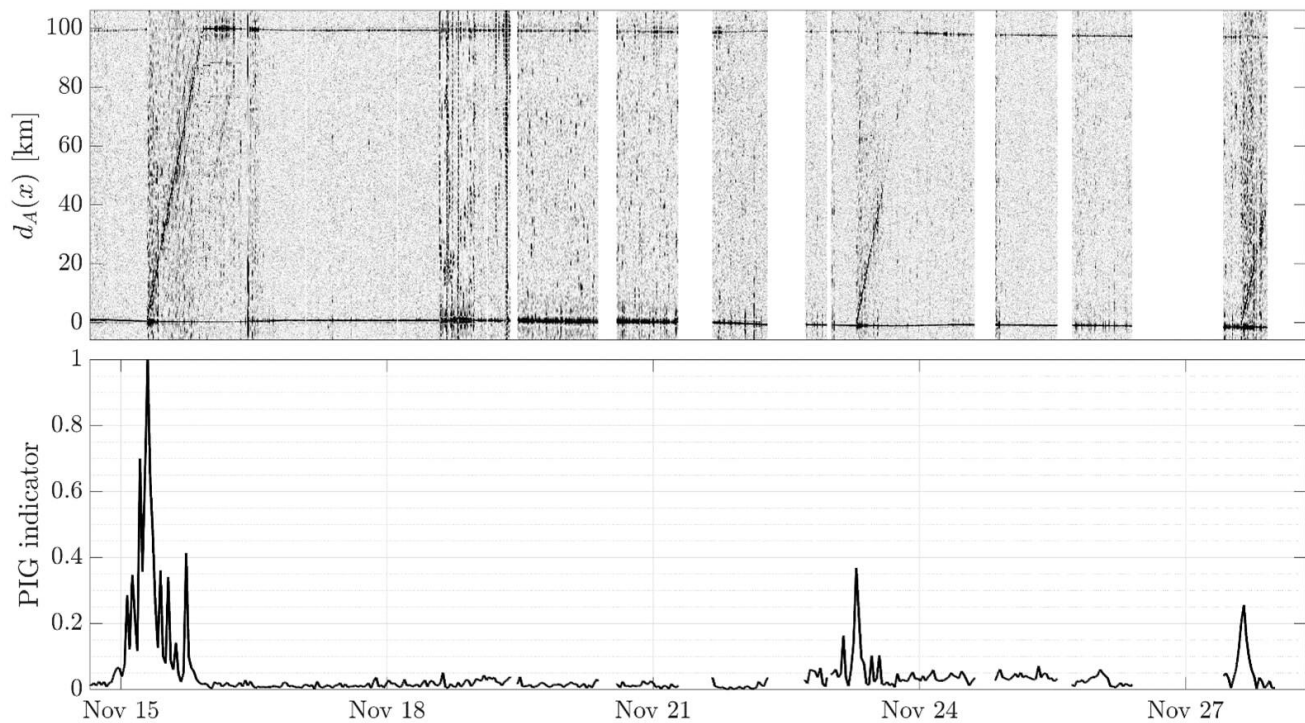

Fig. 6—Short-term (gray line) and long-term (magenta line) specific attenuation for the $\overline{AC}$ pipe segment. The main PIG campaigns (from an attenuation perspective) performed on the pipeline have been highlighted with black vertical bars.

$$d_A\left(x\right) = v \cdot \tau_A\left(x\right) \quad \left[\mathrm{km}\right]. \tag{2}$$

At first sight, two horizontal dark lines located at the cross-correlation distances $d_A\left(x\right) = 100$ km and $d_A\left(x\right) = 0$ km can be noticed: The former is related to the pressure transients that originate from Station C and propagate to the opposite line end, while the latter corresponds instead to the physiological propagation delay of the acoustic waves emitted by the pump located at Station A that reach Station C. Both curves present slight upward and downward curvatures, as the velocity of sound within the pipe is not constant over time: Temperature variations and changes in the composition of the flowing product are mainly responsible for such irregularities (Creek et al. 1999). In addition, several white rectangular areas of the images are present, corresponding to all the data points that have been ruled out as a result of the processing flow described in section Data Processing; in those circumstances, $R_{\overline{AC}}$ cannot be computed and it is automatically set to a null value.

On 15 November at approximately 07:00, a pipeline cleaning operation is initiated, as the PIG departs from Station A to reach the end of the line (Station C) about 16 hours later. The event is observable in the cross-correlation map, as a slant dark line originating from $d_A\left(x\right) = 0$ km and gradually increasing toward $d_A\left(x\right) = 100$ km, for increasing time; this phenomenon can be interpreted as the position of an acoustic source (e.g., the PIG) that is traveling inside the pipes. In the same image, two additional slant lines can also be noticed; they are representative of further runs, respectively, starting on 23 November at 07:00 and on 27 November at 15:00.

Fig. 7—Cross-correlation panel (top), describing the position of several PIGs along the pipeline as a function of time and distance from Station A, and corresponding PIG indicator (bottom), highlighting the start of PIG runs.

Another utility provided by the PIG detection tool consists in the possibility of computing the speed of the moving inspection gauge. Once two consecutive timesteps t_1 and t_2 of a particular run have been identified, the instantaneous PIG velocity v_{PIG} is obtained as:

$$v_{\text{PIG}} = 1000 \frac{d_{\overline{AC}}}{t_2 - t_1} \quad [\text{m/s}]. \tag{3}$$

The bottom plot of **Fig. 7** represents a PIG indicator; namely, it highlights the presence of a coherent cross-correlation peak along a tilted line in the top image of **Fig. 6**. As stated before, this line is the time-pipe coordinate mark of the traveling PIG, and the slope of the line gives the velocity of the gauge. The PIG indicator index $PII(t)$ is computed by summing the data along a range of "realistic" PIG displacement velocities, like in a Hough/Radon transform processing (Illingworth and Kittler 1988).

From the PIG velocity range (v_1, v_2), one can obtain the slope range (m_1, m_2) of a tilted line of type $d_a = mt$:

$$PII(t) = \int_{m_1}^{m_2} dm \frac{\left(\int_{x=mt} d_a(x,t)dx\right)^2}{\int_{x=mt} d_a^2(x,t)dx}. \tag{4}$$

Design of the Inner Deposits Predictor. As stated at the beginning of the paper, the goal of this work consists in the development of an online data-driven procedure that, starting from the short-term attenuation time series (gray curve in **Fig. 6**), can predict the occlusion level of pipeline internals. We have observed a strong correlation with the specific attenuation values; however, we still need to provide a clear and unambiguous definition that numerically quantifies the concept of pipe occlusion. This operation becomes necessary to entirely formalize the problem within a proper machine learning context. For this purpose, the first step consists in defining which type of learning task needs to be solved. We have opted for supervised regression for two main reasons: First, a pipe gradually clogs up because of multiple factors that continuously evolve over the course of several months (e.g., buildup of wax deposits, debris, etc.); as a consequence, classification algorithms are not suitable for this kind of prediction, as they provide discrete outputs (e.g., binary labels, such as clean pipe and clogged pipe) and disregard any intermediate stage; lastly, by predicting the value of a continuous variable through regression, we can express such a variable as an occlusion indicator that can be easily understood by nonexperts in the field. For instance, an automated system can be set up such that if the amount of inner deposits is above a certain threshold, a cleanup campaign is consequently triggered in the pipeline.

Employing supervised learning techniques requires having labeled data at disposal, which are rarely available in pipeline transportation systems (Lygren et al. 2019). To overcome this issue, we have manually built the target function $y_{\overline{ij}}$ to be learned by the supervised regressor and expressed it as a numeric variable ranged between zero and unity. For a given pipeline segment $\overline{ij}$ between two stations i and j, $y_{\overline{ij}}$ is defined as:

$$y_{\overline{ij}} = f\left(\alpha_{\overline{ij},\text{long}}\right), \tag{5}$$

where f is a mapping function of the long-term specific attenuation $\alpha_{\overline{ij},\text{long}}$ between stations i and j (e.g., the magenta curve displayed in **Fig. 6**), which rescales the data in a range comprised between zero and unity. The result of such an operation is displayed in **Fig. 8,** where the time series of $y_{\overline{AC}}$ (corresponding to the $\overline{AC}$ segment) is represented. To make the representation more straightforward, we have
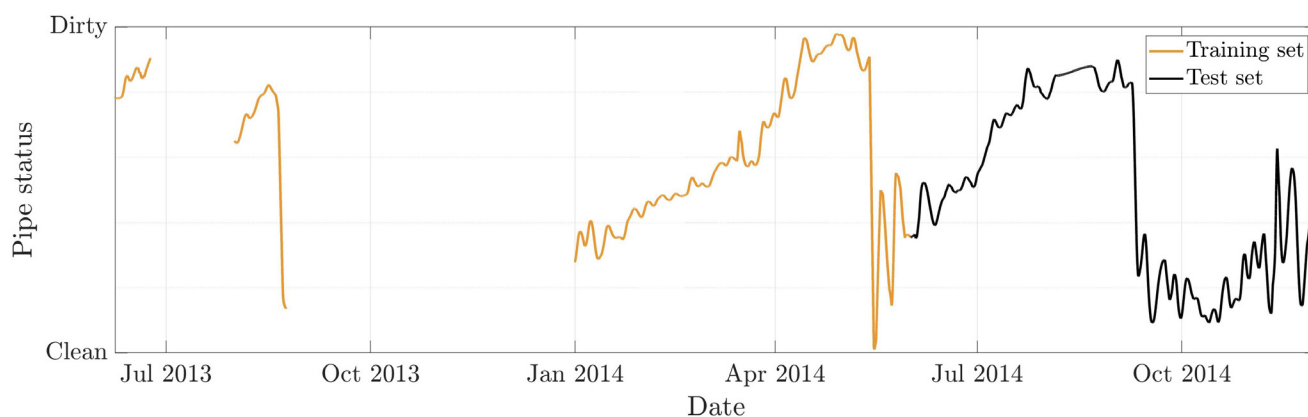
Fig. 8—Target function $y_{\overline{AC}}$ for the line section $\overline{AC}$.

substituted the numeric values displayed on the vertical axis with their qualitative interpretation (e.g., 0 and 1, respectively, translate into clean and dirty). Lastly, the plotted curve presents some gaps, which correspond to missing attenuation samples; in all those circumstances, the machine learning algorithm cannot be either trained or tested.

The learning task is performed by an extremely randomized trees regressor (ERTR) (Geurts et al. 2006), which is a supervised meta-estimator trained to fit several decision tree regressors (DTRs) and to provide a numerical output that is the average of their predictions. In extremely randomized trees, each DTR is created by introducing randomness during their generation phase. ERTRs share the same working principle as standard DTRs, in which a model is fit to the training data based on a set of intuitive decision rules (e.g., if/else statements) (Quinlan 1986), which are directly derived from the input features; such a model can then predict a numerical quantity that is a nonlinear function of the input characteristics. In our case, we have designed an ERTR to provide automatically, as output, a continuous random variable ranged between zero and unity.

Compared with a DTR, implementing an ERTR can be more advantageous for several reasons. First of all, ERTRs do not exhibit the high-variance issues affecting DTRs (Dietterich and Bae 1995), which are usually associated with overfitting a model to the data: As said, the variance of ERTRs is in fact reduced by averaging the estimates provided by several DTRs. Second, they do not favor features having high cardinality (namely, with several unique values) (Louppe 2014), which may become problematic when computing statistical indicators from continuous random variables (e.g., time series data). Lastly, they also provide information, by means of relative rank assessment, on which features contribute the most to the final prediction (Louppe 2014).

Training a supervised regression algorithm requires the evaluation of two quantities—an $m \times N$ matrix of features X and a $m \times 1$ target vector y, where m and N, respectively, correspond to the number of training examples and input characteristics. As previously discussed, y has been computed using (5); each row of X consists instead in a set of statistical indicators, computed over nine different rolling and causal windows, ranging from 8 hours up to 7 days. More precisely, we have chosen to evaluate the mean, minimum, and maximum values of the short-term attenuation, thus resulting in a final set of 27 features. So, for each of the m input examples, the ERTR is fed with a 1×27 feature vector x_k and a 1×1 target scalar y_k.

The ERTR model has been trained using data from $\overline{AC}$ line segment from 1 June 2013 to 31 May 2014 included, whereas the testing phase has been performed on the same line section from 1 June 2014 to 1 December 2014; this results in an even split between the training ($y_{\overline{AC},\text{train}}$, orange line in **Fig. 8**) and the test ($y_{\overline{AC},\text{test}}$, black line in **Fig. 8**) sets, because several months of 2013 are characterized by unavailable data points.

Results and Discussion. The prediction accuracy of the model has been assessed by evaluating the root mean squared value RMS$\{\epsilon\}$ of the estimation error $\epsilon = \hat{y} - y$ between the estimated pigging probability $\hat{y}$ and the true target vector y, and by computing the coefficient of determination (denoted as R^2 score). RMS$\{\epsilon\}$ is defined as follows:
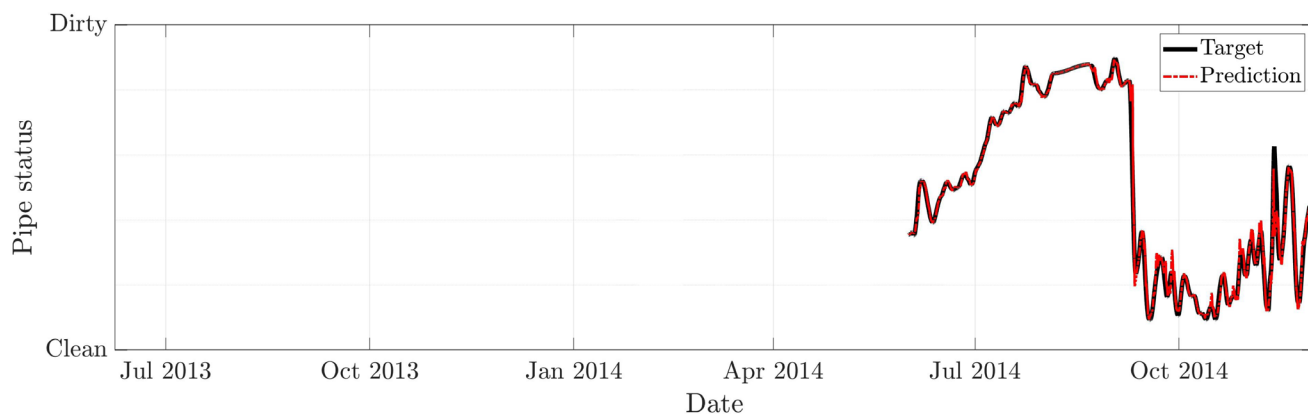

Fig. 9—True (black line) and predicted (red line) values of the target function for the $\overline{AC}$ line section.

$$\text{RMS}\{\epsilon\} = \sqrt{\frac{1}{n}\epsilon^T\epsilon},$$

(6)

where n is the length of ϵ and ϵ^T corresponds to its transpose. Because the target function y had been transformed to be bound between zero and unity, one can also derive the percentage prediction accuracy as $100 \cdot (1 - \text{RMS}\{\epsilon\})$.
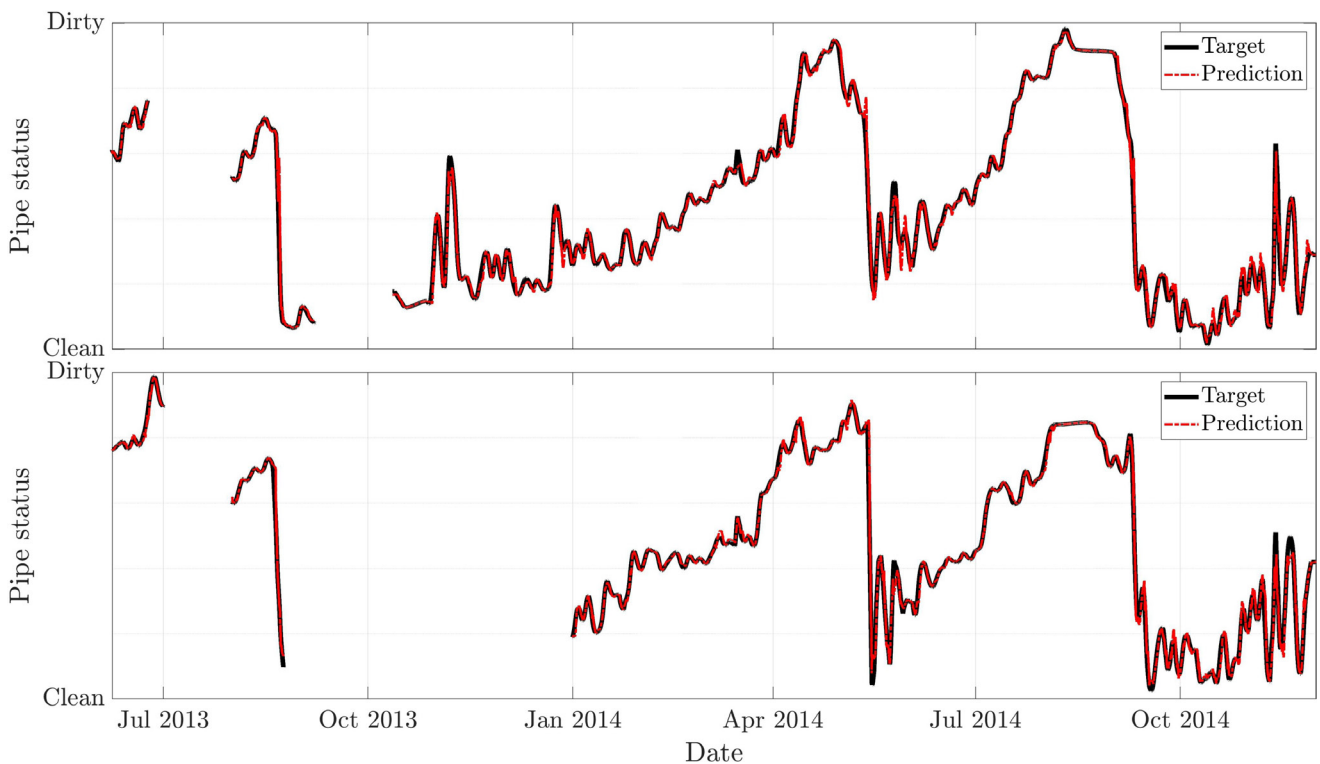
The coefficient of determination, instead, is a widely used goodness of fit indicator in regression problems and measures how precisely unseen data points are going to be predicted by the model. Such a coefficient is ranged between zero and unity: In the former case, the model performs poorly because it always predicts the expected value of y (denoted with $\bar{y}$); in the latter instance, the model perfectly explains the data. R^2 is expressed as:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}\epsilon_k^2}{\sum_{k=1}^{n}(y_k-\bar{y})^2},$$

(7)

where $\epsilon_k^2 = (\hat{y}_k - y_k)^2$ corresponds to the squared estimation error between the kth prediction $\hat{y}_k$ and its target value y_k, while n represents the number of samples in the validation set.

With regard to the test set considered ($y_{\overline{AC},\text{test}}$, black line in **Fig. 8**), we have attained values of RMS $\{\epsilon_{\overline{AC},\text{test}}\}$, prediction accuracy, and R^2 score, respectively, equal to 0.0261, 97.39%, and 0.9906. As a reference, **Fig. 9** graphically compares the values of $y_{\overline{AC},\text{test}}$ (**Fig. 9**, black line) with the predictions ($\hat{y}_{\overline{AC},\text{test}}$, red curve in **Fig. 9**): The performance is satisfactory, as displayed by a good agreement between the two curves.

Once the robustness of the model has been verified on data belonging to the same distribution of the training set, its generalization capabilities must be assessed on additional unseen data. For this purpose, we have used the dynamic pressure measurements collected at Station B to perform further testing of the model on two additional line sections, labeled $\overline{AB}$ and $\overline{BC}$; they, respectively, correspond to the pipeline lengths connecting Station A with Station B and Station B with Station C. Compared with the $\overline{AC}$ case, the entire sample set can now be used for testing, as the model does not require additional training. **Fig. 10** graphically compares the true (black line) and the predicted (red curves) values of the target functions for the $\overline{AB}$ (**Fig. 10**, top) and $\overline{BC}$ (**Fig. 10**, bottom) segments, while **Table 2** summarizes the values of RMS $\{\epsilon\}$, prediction accuracy, and R^2 score for the three pipeline lengths $\overline{AC}$, $\overline{AB}$, and $\overline{BC}$. Once again, the results are quite satisfactory, as the attained accuracy level is greater than 97% in all three cases. In addition, testing measurements collected along different pipeline subsections allow one to determine which portion of the conduit is subject to the highest occlusion levels among the others: For instance, from July 2014 to September 2014, the $\overline{AB}$ segment is more affected by internal deposits than $\overline{BC}$, because the probability indicator of the former is higher than the latter (as displayed in **Fig. 10**).

Fig. 10—True (black line) and predicted (red line) values of the target function for the $\overline{AB}$ (top) and $\overline{BC}$ (bottom) line segments.

Table 2—Performance metrics for the three line segments.

Line Section	RMS $\{\epsilon\}$	Accuracy	R^2
$\overline{AC}$	0.0261	97.39%	0.9906
$\overline{AB}$	0.0196	98.04%	0.9944
$\overline{BC}$	0.0197	98.03%	0.9937

Table 2—Performance metrics for the three line segments.

As a last consideration, the proposed model can potentially be deployed for real-time applications: The input features x_k fed to the ERTR are only dependent on the past history (with respect to the current sample at time k), because they are obtained from moving statistics computed over causal windows, and the evaluation of the instantaneous target y_k can be neglected by accepting the prediction $\hat{y}_k$ to be temporarily unsupervised (we recall that y_k is a function of the long-term attenuation, which has been expressed as noncausal, 1 week moving average of its short-term correspondent). This tradeoff is still acceptable, as it would simply introduce a delay of 3.5 days in the evaluation of the accuracy metrics described at the beginning of this section; the predicted output $\hat{y}_k$, instead, would still be provided instantaneously. Moreover, performance assessment becomes even less urgent whenever the model reliability has already been validated on a sufficiently large data set (e.g., several months or years of historical data).

Conclusion

This paper presents a data-driven methodology to automatically monitor the inner deposits in crude oil transportation pipelines. The proposed solution makes use of standard pressure measurements, collected in two different locations along the pipeline, which are reprocessed and fed to a nonlinear, supervised regression algorithm (ERTR). The latter has been designed to output probability measures that numerically quantify the level of debris within the pipe itself. The data-driven machine learning model has been successfully applied to the vibroacoustic data collected from a crude oil pipeline; its performance has been assessed in terms of prediction accuracy and coefficient of determination, achieving scores, respectively, greater than 97% and 0.99 for all three test sets considered over 18 months. Results obtained so far show the possibility of predicting and tracking the occlusion levels of the entire pipeline and of individual pipe sections; such capabilities prove to be advantageous in the context of planning optimal predictive maintenance strategies, as cleanup campaigns can be triggered only when necessary and on the mostly clogged pipeline sections. In addition, the trained ERTR is potentially employable for real-time integrity assessment applications, thus enabling the opportunity for online monitoring.

Future work includes an additional validation phase on other oil and gas transportation systems, testing on multiphase fluids in the upstream scenario and the definition of optimal threshold criteria which would trigger a new pipeline inspection and operative campaign. We also hypothesize that the attenuation of acoustic waves could potentially be measured from the outside, namely, on the pipe shell by using acceleration sensors or strain gauges—a major advantage of such an implementation is the elimination of internal sensing systems (e.g., temperature or flow rate sensors, etc.), which typically require an inspection chamber, thus guaranteeing additional freedom and convenience in the arrangement of the monitoring setup. This aspect will however be studied and tested more thoroughly in future experiments.

Acknowledgment

References

Adeyanju, O. A. and Oyekunle, L. O. 2019. Experimental Study of Water-in-Oil Emulsion Flow on Wax Deposition in Subsea Pipelines. *J Pet Sci Eng* **182**: 106294. https://doi.org/10.1016/j.petrol.2019.106294.

Agarwal, J. R., Torres, C. F., and Shah, S. 2021. Development of Dimensionless Parameters and Groups of Heat and Mass Transfer to Predict Wax Deposition in Crude Oil Pipelines. *ACS Omega* **6** (16): 10578–10591. https://doi.org/10.1021/acsomega.0c05966.

Alnaimat, F. and Ziauddin, M. 2020. Wax Deposition and Prediction in Petroleum Pipelines. *J Pet Sci Eng* **184**: 106385. https://doi.org/10.1016/j.petrol.2019.106385.

Bernasconi, G., Del Giudice, S., and Giunta, G. 2014. Advanced Real Time and Long Term Monitoring of Transportation Pipelines. Paper presented at the ASME 2014 International Mechanical Engineering Congress and Exposition, Montreal, Quebec, Canada, 14–20 November. IMECE2014-36872. https://doi.org/10.1115/IMECE2014-36872.

Bernasconi, G. and Giunta, G. 2020. Acoustic Detection and Tracking of a Pipeline Inspection Gauge. *J Pet Sci Eng* **194**: 107549. https://doi.org/10.1016/j.petrol.2020.107549.

Blackstock, D. T. and Atchley, A. A. 2001. Fundamentals of Physical Acoustics. *J Acoust Soc Am* **109** (4): 1274–1276. https://doi.org/10.1121/1.1354982.

Chen, S., Chen, Y., Jiang, J. et al. 2021. Wax Deposition Model of Heavy Oil in Pipeline Transportation. In *E3S Web Conference*, Vol. **260**, 01003. https://doi.org/10.1051/e3sconf/202126001003.

Chi, Y., Sarica, C., and Daraboina, N. 2019. Experimental Investigation of Two-Phase Gas-Oil Stratified Flow Wax Deposition in Pipeline. *Fuel* **247**: 113–125. https://doi.org/10.1016/j.fuel.2019.03.032.

Creek, J. L., Lund, H. J., Brill, J. P. et al. 1999. Wax Deposition in Single Phase Flow. *Fluid Phase Equilib* **158–160**: 801–811. https://doi.org/10.1016/S0378-3812(99)00106-5.

Dietterich, T. G. and Bae, K. E. 1995. Machine Learning Bias, Statistical Bias, and Statistical Variance of Decision Tree Algorithms. Technical report. Department of Computer Science, Oregon State University.

Geurts, P., Ernst, D., and Wehenkel, L. 2006. Extremely Randomized Trees. *Mach Learn* **63** (1): 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Giacchetta, G., Marchetti, B., Leporini, M. et al. 2019. Pipeline Wax Deposition Modeling: A Sensitivity Study on Two Commercial Software. *Petroleum* **5** (2): 206–213. https://doi.org/10.1016/j.petlm.2017.12.007.

Giro, R. A., Bernasconi, G., Giunta, G. et al. 2021. A Data-Driven Pipeline Pressure Procedure for Remote Monitoring of Centrifugal Pumps. *J Pet Sci Eng* **205**: 108845. https://doi.org/10.1016/j.petrol.2021.108845.

Giunta, G. and Bernasconi, G. 2019. Method and System for Continuous Remote Monitoring of the Integrity of Pressurized Pipelines and Properties of the Fluids Transported. US Patent No 10401254B2.

Giunta, G., Bernasconi, G., Giro, R. A. et al. 2020. Digital Transformation of Historical Data for Advanced Predictive Maintenance. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 9–12 November. SPE-202906-MS. https://doi.org/10.2118/202906-MS.

Giunta, G., Morrea, S., Gabbassov, R. et al. 2016. Performance of Vibroacoustic Technology for Pipeline Leak Detection. In *American Society of Mechanical Engineers, Pressure Vessels and Piping Division (Publication) PVP*. https://doi.org/10.1115/OMAE2016-54181.

Guozhong, Z. and Gang, L. 2010. Study on the Wax Deposition of Waxy Crude in Pipelines and Its Application. *J Pet Sci Eng* **70** (1–2): 1–9. https://doi.org/10.1016/j.petrol.2008.11.003.

Halstensen, M., Arvoh, B. K., Amundsen, L. et al. 2013. Online Estimation of Wax Deposition Thickness in Single-Phase Sub-Sea Pipelines Based on Acoustic Chemometrics: A Feasibility Study. *Fuel* **105**: 718–727. https://doi.org/10.1016/j.fuel.2012.10.004.

Huang, Q. and Ma, J. 2008. Prediction of Wax Deposition of Crude Using Statistical and Neural Network Methods. Paper presented at the 2008 7th International Pipeline Conference, Calgary, Alberta, Canada, 29 September–3 October. IPC2008-64225. https://doi.org/10.1115/IPC2008-64225.

Huang, Q., Wang, W., Li, W. et al. 2017. A Pigging Model for Wax Removal in Pipes. *SPE Prod & Oper* **32** (4): 469–475. SPE-181560-PA. https://doi.org/10.2118/181560-PA.

Illingworth, J. and Kittler, J. 1988. A Survey of the Hough Transform. *Comput Vision Graph Image Process* **44** (1): 87–116. https://doi.org/10.1016/S0734-189X(88)80033-1.

Ito, S., Tanaka, Y., Hazuku, T. et al. 2021. Wax Thickness and Distribution Monitoring Inside Petroleum Pipes Based on External Temperature Measurements. *ACS Omega* **6** (8): 5310–5317. https://doi.org/10.1021/acsomega.0c05415.

Jalalnezhad, M. J. and Kamali, V. 2015. Development of an Intelligent Model for Wax Deposition in Oil Pipeline. *J Petrol Explor Prod Technol* **6** (1): 129–133. https://doi.org/10.1007/s13202-015-0160-3.

Kamari, A., Khaksar-Manshad, A., Gharagheizi, F. et al. 2013. Robust Model for the Determination of Wax Deposition in Oil Systems. *Ind Eng Chem Res* **52** (44): 15664–15672. https://doi.org/10.1021/ie402462q.

Kamari, A., Mohammadi, A. H., Bahadori, A. et al. 2014. A Reliable Model for Estimating the Wax Deposition Rate During Crude Oil Production and Processing. *Pet Sci Technol* **32** (23): 2837–2844. https://doi.org/10.1080/10916466.2014.919007.

Leporini, M., Terenzi, A., Marchetti, B. et al. 2019. Experiences in Numerical Simulation of Wax Deposition in Oil and Multiphase Pipelines: Theory versus Reality. *J Pet Sci Eng* **174**: 997–1008. https://doi.org/10.1016/j.petrol.2018.11.087.

Li, W., Huang, Q., Wang, W. et al. 2020. Advances and Future Challenges of Wax Removal in Pipeline Pigging Operations on Crude Oil Transportation Systems. *Energy Technol* **8** (6). https://doi.org/10.1002/ente.201901412.

Li, W., Huang, Q., Wang, W. et al. 2018. Study on Wax Removal During Pipeline-Pigging Operations. *SPE Prod & Oper* **34** (1): 216–231. SPE-194010-PA. https://doi.org/10.2118/194010-PA.

Li, W., Li, H., Da, H. et al. 2021. Influence of Pour Point Depressants (PPDs) on Wax Deposition: A Study on Wax Deposit Characteristics and Pipeline Pigging. *Fuel Process Technol* **217**: 106817. https://doi.org/10.1016/j.fuproc.2021.106817.

Lock Sow Mei, I., Ismail, I., Shafquet, A. et al. 2015. Real-Time Monitoring and Measurement of Wax Deposition in Pipelines via Non-Invasive Electrical Capacitance Tomography. *Meas Sci Technol* **27** (2). https://doi.org/10.1088/0957-0233/27/2/025403.

Louppe, G. 2014. *Understanding Random Forests: From Theory to Practice*. PhD Dissertation, University of Liège, Liège, Belgium. https://doi.org/10.13140/2.1.1570.5928.

Lygren, S., Piantanida, M., and Amendola, A. 2019. Unsupervised, Deep Learning-Based Detection of Failures in Industrial Equipments: The Future of Predictive Maintenance. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 11–14 November. SPE-197629-MS. https://doi.org/10.2118/197629-MS.

McAllister, E. W. 2013. *Pipeline Rules of Thumb Handbook*, eight edition. Houston, Texas, USA: Gulf Professional Publishing. https://doi.org/10.1016/C2013-0-00277-0.

Menad, N. A., Jahanbani Ghahfarokhi, A., and Shang Wui Ng, C. 2021. Predicting Wax Deposition Using Robust Machine Learning Techniques. *Petroleum*(in press; available online 4 July 2021). https://doi.org/10.1016/j.petlm.2021.07.005.

Modesty Kelechukwu, E., Said Al-Salim, H., and Saadi, A. 2013. Prediction of Wax Deposition Problems of Hydrocarbon Production System. *J Pet Sci Eng* **108**: 128–136. https://doi.org/10.1016/j.petrol.2012.11.008.

Mwendapole Lonje, B. and Liu, G. 2021. Review of Wax Sedimentations Prediction Models for Crude-Oil Transportation Pipelines. *Pet Res*. https://doi.org/10.1016/j.ptlrs.2021.09.005.

Obanijesu, E. O. and Omidiora, E. O. 2008. Artificial Neural Network's Prediction of Wax Deposition Potential of Nigerian Crude Oil for Pipeline Safety. *Pet Sci Technol* **26** (16): 1977–1991. https://doi.org/10.1080/10916460701399485.

Obaseki, M. and Elijah, P. T. 2021. Dynamic Modeling and Prediction of Wax Deposition Thickness in Crude Oil Pipelines. *J King Saud Univ Eng Sci* **33** (6): 437–445. https://doi.org/10.1016/j.jksues.2020.05.003.

Obinichi, N., Nwachukwu Okpala, A., and Johnnie Tuaweri, T. 2021. Influence of Temperature on Wax Deposit on Corrosion of Crude Oil Pipeline. *Am J Mech Mater Eng* **5** (2): 29–34. https://doi.org/10.11648/j.ajmme.20210502.12.

Olajire, A. A. 2021. Review of Wax Deposition in Subsea Oil Pipeline Systems and Mitigation Technologies in the Petroleum Industry. *Chem Eng J Adv* **6**: 100104. https://doi.org/10.1016/j.ceja.2021.100104.

Quinlan, J. R. 1986. Induction of Decision Trees. *Mach Learn* **1** (1): 81–106. https://doi.org/10.1007/BF00116251.

Shasha, H. and Qiyu, H. 2014. Research on Wax Deposition in the Pipeline Without Pigging for a Long Time. *Pet Sci Technol* **32** (3): 316–323. https://doi.org/10.1080/10916466.2011.578090.

Sousa, A. M., Pereira, M. J., Matos, H. A. et al. 2021. Planning Pipeline Pigging Operations with Predictive Maintenance. In *E3S Web Conference*, Vol. 266, 01017. https://doi.org/10.1051/e3sconf/202126601017.

Sun, D., Zhu, Z., Hu, Z. et al. 2020. Experimental and Theoretical Study on Wax Deposition and the Application on a Heat Insulated Crude Oil Pipeline in Northeast China. *Oil Gas Sci Technol* **75**: 3. https://doi.org/10.2516/ogst/2019064.

Theyab, M. A. and Diaz, P. 2016. Experimental Study of Wax Deposition in Pipeline – Effect of Inhibitor and Spiral Flow. *Int J Smart Grid Clean Energy* **5** (3): 174–181. https://doi.org/10.12720/sgce.5.3.174-181.

Van der Geest, C., Guersoni, V. C. B., Merino-Garcia, D. et al. 2018. Wax Deposition Experiment with Highly Paraffinic Crude Oil in Laminar Single-Phase Flow Unpredictable by Molecular Diffusion Mechanism. *Energy Fuels* **32** (3): 3406–3419. https://doi.org/10.1021/acs.energyfuels.8b00269.

Van der Geest, C., Melchuna, A., Bizarre, L. et al. 2021. Critical Review on Wax Deposition in Single-Phase Flow. *Fuel* **293**: 120358. https://doi.org/10.1016/j.fuel.2021.120358.

Wang, W. and Huang, Q. 2014. Prediction for Wax Deposition in Oil Pipelines Validated by Field Pigging. *J Energy Inst* **87** (3): 196–207. https://doi.org/10.1016/j.joei.2014.03.013.

Wang, W., Huang, Q., Huang, J. et al. 2014a. Study of Paraffin Wax Deposition in Seasonally Pigged Pipelines. *Chem Technol Fuels Oils* **50** (1): 39–50. https://doi.org/10.1007/s10553-014-0488-2.

Wang, W., Huang, Q., Li, S. et al. 2014b. Identifying Optimal Pigging Frequency for Oil Pipelines Subject to Non-Uniform Wax Deposition Distribution. Paper presented at the 2014 10th International Pipeline Conference, Calgary, Alberta, Canada, 29 September–3 October. IPC2014-33064. https://doi.org/10.1115/IPC2014-33064.

Xie, Y., Chen, D., and Mai, F. 2018. Economic Pigging Cycles for Low-Throughput Pipelines. *Adv Mech Eng* **10** (11). https://doi.org/10.1177/1687814018811198.

Xie, Y. and Xing, Y. 2017. A Prediction Method for the Wax Deposition Rate Based on A Radial Basis Function Neural Network. *Petroleum* **3** (2): 237–241. https://doi.org/10.1016/j.petlm.2016.08.003.

Yao, B., Zhao, D., Zhang, Z. et al. 2021. Safety Study on Wax Deposition in Crude Oil Pipeline. *Processes* **9** (9): 1572. https://doi.org/10.3390/pr9091572.

Zougari, M. I. 2010. Shear Driven Crude Oil Wax Deposition Evaluation. *J Pet Sci Eng* **70** (1–2): 28–34. https://doi.org/10.1016/j.petrol.2009.01.011.