Wavelet transform processing in detecting failures in offshore well production

Priscila Esposte Coutinho¹,*, Larissa Haringer Martins da Silveira¹, Marcio Cataldi², Fabiana Rodrigues Leta², Antônio Orestes de Salvo Castro³, Cláudio Benevenuto de Campos Lima⁴, Gilson Brito Alves Lima²

¹ Postgraduate students, Universidade Federal Fluminense – UFF, Rio de Janeiro, Brazil
² Teacher and Researcher, Universidade Federal Fluminense – UFF, Rio de Janeiro, Brazil
³ Teacher and Researcher, Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro, Brazil
⁴ Manager, Petróleo Brasileiro S.A – Petrobras, Rio de Janeiro, Brazil

*Autor para correspondência, E-mail: priscila_esposte@id.uff.br

Received: 07 June 2022 | Accepted: 14 July 2022 | Published online: 31 August 2022

Abstract: Brazil has a significant offshore oil production, which dates back to the late 1960s and is currently focused on exploring pre-salt reservoirs. The drilling technology Petrobras uses is considered a world standard: in 2020, it allowed offshore production to reach 97% of the country’s total oil production. During the process, however, unwanted events, and even operational failures may occur, which are capable of significant damage. Thus, failure detection is extremely important to prevent production losses or delays, to reduce costs and to avoid accidents. This study uses a real, public database on offshore production, and proposes using wavelet transforms to detect production failures. With the technique, we pinpointed which time intervals between measurements showed relevant variability, and then clustered the data, according to mobile averages, to shrink the record number. Using wavelet transforms, we analyzed which variables could be used as predictors of production failures and identified the temperature read by the Temperature and Pressure Transducer sensor (T-TPT) and the pressure at the Production Choke sensor (P-PCK) as possible predictor variables. We also observed the creation of a filtered series, averaged from the original data series, which maintained its variability, showing the viability of record regrouping in shorter series.

Keywords: Offshore wells, oil production, wavelet transform, predictor variables, time series.
1. Introduction

Offshore oil production in Brazil dates back to the late 1960s, with the discovery of the Guaricema field in Sergipe. It greatly expanded over the following decade, when the Campos Basin reservoirs started operating. The region quickly became the country’s main oil producer up until this century’s first decade when Santos Basin became the leading producer with the start of pre-salt exploration (D’Almeida, 2015). In 2020, offshore production hit 2,485 million barrels per day (b/d), almost 97% of the total production (ANP 2021). Petrobras’ production is considered a world standard due to its technology and results. In building the wells, the company uses sophisticated equipment with a wide set of sensors along the drilling strings and production lines that reach the seabed (Ortiz Neto and Costa, 2007).

In drilling the wells, more than 50 parameters are continuously monitored: drill weight, pipe rotation, flow, pressure and torque, among others. (Marques et al., 2019). Subsequently, the actual extraction of the fluids in the reservoirs, several unwanted events that would cause great production losses may occur. Among those cited in Vargas et al. (2019), who used a public dataset of real undesirable events, are the spurious closure of the DHSV valve (down hole safety valve), flow instability, restriction of the PCK valve (choke) and hydrate occurrence in the production line.

Detecting these events, therefore, is of utmost importance in preventing production losses and delays (lost profits), reducing maintenance and intervention (operational costs) and avoiding accidents and their possible consequences. Artificial and real databases are used to develop and enhance these detection techniques, through which all sorts of diagnostic algorithms can be elaborated and tested.

Nonetheless, an important problem regarding the treatment and use of data from oil-producing wells is that, since they are usually measured at such a high frequency, classification and/or prediction models may incur in scalability issues (Martí et al., 2015; Takei et al., 2010; Vargas et al., 2019). To solve this issue, we propose applying wavelet transformation to the database created by Vargas et al. (2019), which contains measurements every second or minute.

Wavelets are functions that can represent series or other functions at different scales or resolutions, and are often used in signal processing. They are, therefore, tools capable of simultaneously locating a signal in space and frequency, providing more compact representations than the original domain (Hammond et al. 2011).

Altogether, they can be understood as a more efficient way of observing a time series regularity and singularity; working in a way similar to variable-sized sliding windows, in which energy peaks are equivalent to behavioral changes (Ray et al., 2011). The technique, therefore, analyzes raw characteristics and refines the data, since it identifies the main variability frequencies in a series and its evolution over time (Torrence and Compo, 1998).

Li et al. (2016) also highlight the possibility of using wavelet transforms to reduce the time series dimensionality, allowing for a classification accuracy similar to the use of original, uncompressed data. The authors also point out its ability to implicitly smooth out data, making it more efficient in classifying time series than explicit techniques.

In the oil sector, wavelet transforms have already been used in different approaches. Naccache (2011), Aguiar-Conraria and Soares (2010), and Reboredo and Rivera-Castro (2014), for instance, performed multiresolution wavelet analyses to measure variations in oil prices under different time scales; whereas Korovin and Khisamutdinov (2014) used a hybrid wavelet method to identify oil pump malfunctions. Layouni et al. (2017), on the other hand, used wavelets associated to neural networks to detect metal loss in pipelines.

Wavelet transforms also allowed Asgarian et al. (2016) to detect damages to offshore fixed platforms, while Zadkarami et al. (2016) proposed a method for assessing leakage presence, possible location, and severity in hydrocarbon ducts using wavelet transforms combined with neural network (multi-layer perceptron) techniques as classifiers for the detection system and fault isolation.

In our study, wavelet analysis served two objectives. The first one was to look for the time intervals between measurements that actually presented the relevant variability regarding Class change or failure occurrence. This type of analysis aimed to observe if the measurements are actually performed at the frequency indicated by the sensor manufacturer; assessing if the equipment inertia is in accordance to its specification; or if the measurement frequency is simply higher than the inertia of the measured quantity. In general, this analysis can indicate at which frequencies the measurements begin to present considerable variability, enabling the use of clustering techniques and moving averages to decrease the number of records without affecting data universe variability.
The second objective was to determine the variables that could be used as possible predictors of changes to the data class. This analysis was done by evaluating which variables presented relevant changes in their energy spectrum at the very moment of or before the class change.

This study is therefore, organized into four sections: Section 1 consists in the subject introduction. Section 2 discusses the methodology used; section 3, the analysis and discussion of the results; and, finally, section 4 presents the conclusion.

2. Material and methods

2.1. Mathematical formulation of the Wavelet Transform

Wavelet transforms (WT) refer to a set of wave-shaped functions generated by dilations \((\psi(t) \rightarrow \psi(2t))\) and translations \((\psi(t) \rightarrow \psi(t + 1))\) of a function of a real variable.

This transformation’s mathematical principle is the creation of a new space based on a standard function of finite energy, sometimes called the mother wavelet, obtained by the following expression (Vitorino et al., 2006; Blain and Kayano, 2011), observed in Eq. (1):

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right),
\]

where \(a, b \in \mathbb{R}\) and \(a \neq 0\); \(a\) being the dilation factor and \(b\), the translation factor.

Parameter \(a\) determines the oscillation frequency and wavelet length; and translation parameter \(b\) determines its displacement position. Usually, \(a\) and \(b\) take special values: \(a = 2^{-j}\) and \(b = k2^{-j}\), with \(j\) and \(k \in \mathbb{Z}\).

Factor \(\frac{1}{\sqrt{a}}\) is every daughter wavelet energy normalizing constant; so that, together, they retain the same energy as the main wavelet. The daughter wavelet equation can be expressed by Eq. (2) (Bolzan, 2006):

\[
\psi_{j,k}(t) = \frac{1}{\sqrt{2}} \psi \left( \frac{t-k}{j} \right),
\]

where \(j, k \in \mathbb{R}\) and \(j \neq 0\); \(j\) being the dilation factor and \(k\), the translation factor.

Finally, the wavelet transform in relation to \(\psi\), is expressed by Eq. (3) (Vitorino et al., 2006; Goswami and Chan, 2011):

\[
(W_{\psi}f)(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi \left( \frac{t-b}{a} \right) dt,
\]

where \(a, b \in \mathbb{R}\), and \(\neq 0\).

For this study, we chose the transformation available in the MATLAB 2016 software, i.e., the Morlet Wavelet function; which may be seen as a periodic function whose amplitude is modulated by a Gaussian function (Torrence and Compo, 1998; Vitorino et al., 2006).

2.2 Data treatment

This study analyzed the public, real database describing rare and undesirable events in oil wells, compiled and provided by Vargas et al. (2019). Such data consist of a set of measurements of commonly monitored variables in offshore wells (P-PDG, T-PDG, P-TPT, T-TPT, P-PCK, T-PCK, P-GLCK, T-GLCK, and QGL), obtained at different dates and times, related to eight types of problems and undesirable events that took place in 21 oil wells. Table 1 shows definitions and other information for each of the sensors analyzed.

A class ranging from 1 to 8 was assigned to each of the possible problems. It comprised the transition period between the normal state of the well and the complete establishment of a failure (transient state); values from 101 to 108 were assigned to each of these failures.
Table 1. Information on the sensors used in this study.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Description</th>
<th>Unit of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-PDG</td>
<td>Pressure at the permanent downhole gauge sensor</td>
<td>Kgf/cm²</td>
</tr>
<tr>
<td>T-PDG</td>
<td>Temperature at the permanent downhole gauge sensor</td>
<td>Celsius</td>
</tr>
<tr>
<td>P-TPT</td>
<td>Pressure at the temperature and pressure transducer sensor</td>
<td>Kgf/cm²</td>
</tr>
<tr>
<td>T-TPT</td>
<td>Temperature at the temperature and pressure transducer sensor</td>
<td>Celsius</td>
</tr>
<tr>
<td>P-PCK</td>
<td>Pressure at the production choke sensor</td>
<td>Kgf/cm²</td>
</tr>
<tr>
<td>T-PCK</td>
<td>Temperature at the production choke sensor</td>
<td>Celsius</td>
</tr>
<tr>
<td>P-GLCK</td>
<td>Pressure at the gas lift choke sensor</td>
<td>Kgf/cm²</td>
</tr>
<tr>
<td>T-GLCK</td>
<td>Temperature at the gas lift choke sensor</td>
<td>Celsius</td>
</tr>
<tr>
<td>QGL</td>
<td>Flowrate at the gas lift sensor</td>
<td>m³.day⁻¹</td>
</tr>
</tbody>
</table>

2.3. Wavelet transform application

To analyze oil wells failures via Wavelet Transforms, it was necessary to treat the data so they would meet the analytical tool’s criteria: the exclusion of punctual failures to obtain a continuous series, for example. Files were analyzed along a problem-well axis, prioritizing the observation of each problem verified in each well.

The subsequent step was performing the wavelet transforms using the MATLAB software. We then created figures that would describe the temporal evolution of the variables, as a visual representation of the time series, and the wavelets itself, as a representation of the frequency and the power spectrum of these series. Thus, a pair of figures was created for each variable – one related to the time series description, and another, to the wavelet.

Subsequently, we did the wavelet analysis, having figures of the variable Class as its starting point. Since wavelets are tools capable of representing energy variations in specific frequencies, and even signal discontinuities, we could observe greater energy peaks when the behavior of the well changed, such as from normal to a transient state, and from the latter to the establishment of the failure.

These peaks were marked and reflected to other wavelets, allowing us to observe if each variables’ major energy alterations occurred concomitantly to class changes. This step tried to determine which of the analyzed variables could be potential predictors of well failures, i.e., which would present behavioral changes (greater energy accumulation/better defined peaks) before class changes were observed.

Having analyzed and selected the possible predictors, this study’s first objective was then undertaken: finding time intervals in which relevant variability in the time series occurred. All wavelets were analyzed once again, pinpointing when the first energy peaks actually occurred, and what were their frequencies. For most figures, we observed that the first considerable records took place only after 16 seconds. This underpinned the application of a clustering technique aimed at shrinking the data size of original files without affecting its variability.

With the R software, we then developed a script capable of estimating each variable mean in a 16-second interval, relating the obtained average values to the appropriate indices in columns “class” and “time”. These new time series, i.e., the original series average values, were saved in another file, and new wavelets were created with these mean-filtered data.

This last step aimed to observe whether clustering records into average series would be viable, i.e., if the same structure of signal frequency and energy would be maintained in the filtered, 16-time reduced series, preserving the data variability. This analysis indicates if we could reduce data scalability to use artificial intelligence in failure classification.
Finally, to evaluate the compatibility of variability between the original and the filtered series, we estimated the relative difference of the following metrics: arithmetic mean, standard deviation, variance and the interquartile distance (P75%-P25%).

3. Results and discussion

This study will only address the results related to Problem 8 – Hydrate Formation, in light of its capacity to negatively affect production rates. According to Vargas et al. (2019), hydrates are crystalline compounds formed by water and natural gases under high pressures, and low temperatures. They often occur in oil and gas pipelines. In oil production, their formation can interrupt the flow of a well.

According to the database used, only three wells (19, 20 and 21) exhibited Problem 8; each showing, respectively, 51,199, 15,688 and 24,204 records over time. Thus, due to data volume, we chose to show only the results of Well 21, since its record number lies between the others, and presented less invalid data (about 0.60%).

We will show the file characteristics, the figures related to the behavior of the variable over time, and the wavelet graph. Then, we will present the mean series graphs, thus allowing the comparison of results regarding the preservation of variability in the original data.

3.1. Full series

As for Well 21, time series lacked relevant inconsistencies during data processing: less than two hundred records were suppressed, approximately 0.60% of the total data. We also point out the absence of data on variables T-PCK and T-GLCK, thus totaling a time series of about 24,053 records for each of the five remaining variables.

For hydrate formation (problem 8), classes were categorized numerically: 0 represented the normal state of the well; 108, its transient state, in which it approached failure or operational abnormality due to hydrate formation; and 8, the moment of failure (Vargas et al., 2019).

Regarding the representation of the time variables, Figure 1 shows the time series behavior of each variables of Well 21. The similar behavior of bottom (P-PDG and P-TPT) and surface variables (P-PCK and P-GLCK) is noteworthy. We note two significant class variations: one before 5,000 seconds, and another after 20,000. They indicate the shift from the normal to the transient state, and then to the hydrate formation problem, causing of operational failure.
Wavelets in Figure 2, in turn, show the main energy variations and respective periods/frequencies of each variable. We emphasize that these variations mainly occur up until 5,000 seconds and after 15,000 seconds; whereas between 10,000 and 15,000 seconds low energy is measured, an interval in which no wavelet registers significant peaks.

We also note that the bottom variables P-PDG and P-TPT show higher energy after the change from normal state to failure, suggesting that these may be unreliable predictors for the studied well. The T-TPT variable, however, shows peaks of high energy prior to state change, making it a potential predictor. The surface variables P-PCK and P-GLCK also registered energy peaks (though weaker) before the change to the failure state, especially regarding P-PCK, indicating a potential predictor nature.

After identifying the variables that could be used as predictors of failure caused by hydrate formation, the following step was to try to shrink the time series data size without variability loss. For such purpose, we observed the periods/frequency in which the wavelets of these variables lacked the relevant data variability, i.e., in which they showed insignificant energy values (Power Spectrum).

The analysis was necessary to shrink the size of the time series; since it contained 24,000 measurements in only a small record period. In general, no significant variations were observed in periods shorter than 64 seconds, suggesting that the sensors may have been insufficiently sensitive to show variations within measurements of this interval.

P-PCK wavelet (Figure 2), however, showed small energy variations from 16 seconds onwards; an interval shorter than the 64 seconds observed in other variables. Since this variable was chosen as a potential predictor, we opted to truncate the series into 16-second periods (further discussed in section 4) by estimating the mean of the record within that period, thus eliminating the high-frequency noise that would increase the size of the series sample without bringing information useless for predictor models.
3.2. 16-Second averaged series

After identifying the shorter period that presented energy variation in the wavelets of variables capable of operational failure prognosis and identification, mean values were estimated for this time interval (16 seconds), significantly shrinking the series size. Thus, the record number for each variable changed from around 24,000 to about 1,500.

Regarding the representation of variables in time, Figure 3 shows the behavior of the time series after the creation of a new file with data averaged every 16 seconds. Analysis of the figure shows that filtering the data yielded results very similar to those obtained with the full series in Figure 1.

Once again, we highlight the analogous behavior of bottom (P-PDG and P-TPT) and surface variables (P-PCK and P-GLCK), whose curves ascend over time. The T-TPT variable also behaves similarly as the entire series, in which its curve descended in the interval analyzed. Classes maintained the two relevant variations, suggesting a change from the normal to the transient and then to failure, respectively.
Figure 3. Graphical representation of data on the means of variables P-PDG, P-TPT, T-TPT, P-PCK and P-GLCK of Well 21 regarding problem 8. For each variable, its 16-second time series behavior throughout time and the points of greater variation are shown.

Figure 4 shows wavelets formed from the mean series, in which the main frequencies of each variable can be observed. We point out that, although the energy spectra are less intense if compared to the figures constructed from the unfiltered series, they still behave similarly to those, maintaining their variability structure.

We also highlight that the bottom variables P-PDG and P-TPT still present the highest energy peak after the normal state change, whereas T-TPT shows high energy peaks before the state change, keeping it as a potential predictor. We observed, again, that among the surface variables, energy peaks occur before the change to the transient state, also suggesting them as potential predictors.
Finally, percentage differences were estimated between the mean, standard deviation, variance and interquartile distance of time series of each variable, before and after the application of the filter suggested by the wavelets in Figure 2. These values can be found in Table 2, in which we can observe that, except for the standard deviation and P-PCK variance, all other differences were lower than 2%. The average of all deviations was around 0.7%, i.e., less than 1%, and makes the process of filtering data into smaller series quite feasible.

Table 2. Percentage difference between mean, standard deviation, variance and interquartile distance of time series before and after filter application.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P-PDG</td>
<td>0.0%</td>
<td>0.9%</td>
<td>1.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>P-TPT</td>
<td>0.1%</td>
<td>0.2%</td>
<td>0.5%</td>
<td>0.3%</td>
</tr>
<tr>
<td>T-TPT</td>
<td>-0.2%</td>
<td>0.3%</td>
<td>0.6%</td>
<td>0.8%</td>
</tr>
<tr>
<td>P-PCK</td>
<td>0.4%</td>
<td>2.1%</td>
<td>4.1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>P-GLCK</td>
<td>0.0%</td>
<td>0.3%</td>
<td>0.6%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>
4. Conclusion

This study had two objectives achieved using the wavelet technique, namely: defining possible predictor variables for the hydrate formation problem in oil wells and evaluating the possibility of shrinking the time series – containing measurements every second – without losing its variability.

a) Regarding the first objective, we managed to identify the temperature at the Temperature and Pressure Transducer sensor (T-TPT) and the pressure at the Production Choke sensor (P-PCK) as possible predictor variables, since they exhibited higher energy peaks before the records indicating class change, which suggests that sudden behavioral variations may indicate possible alterations in the state of the well;

b) The surface variables P-PCK and pressure at the Gas Lift Choke (P-GLCK) also showed expressive energy peaks, pointing to a relation with the behavior of the well and making them potential predictors. Thus, the analysis of surface variables should not be completely disregarded when dealing with the problems described in the database, since they may still provide valuable information;

c) Regarding the second objective, we could establish a cut-off period, in which we observed no relevant variability losses in any time series. Once we identified the predictors, we created a new series containing 16-second averaged data, in which the original series variability was maintained. Both original and filtered series had insignificant discrepancies in mean, standard deviation, variance and interquartile difference. This shows that record regrouping is viable in shorter series, if signal frequency is preserved, and suggests that reducing file size has little influence on the quality and robustness of analyses;

d) The application of the wavelet technique to time series of variables of oil exploration, therefore, may be useful in detecting potential predictors of operational failures (in our study, hydrate formation) and in reducing the number of records in these time series without variability losses. Hence, they may represent computational gains, especially when working with larger series, and also help the well control in real time, since the variables behavior may indicate the imminence of a failure, which allows to proceed with alternatives to prevent its complete implementation.

Acknowledgments

The authors gratefully acknowledge UFF – Universidade Federal Fluminense and Petrobras/CENPES – Petróleo Brasileiro S.A. for financial support through the R&D Project no. 21354-6. The authors also state their thanks to Brazilian National Council for Scientific and Technological Development (CNPq - 311484/2020-4) for the financial support.

References


