

Probabilistic Forecasting for Oil Producing Wells Using Seq2seq Augmented Model [†]

Hadeel Affi ^{1,*} , Mohamed Elmahdy ², Motaz El Saban ^{2,3} and Mervat Abu-Elkheir ¹ 

¹ Faculty of Media Engineering and Technology, German University in Cairo, Cairo 11511, Egypt; mervat.abuelkheir@guc.edu.eg

² Raisa Energy LLC, Cairo 11728, Egypt; melmahdy@raisaenergy.com (M.E.); melsaban@raisaenergy.com (M.E.S.)

³ Faculty of Computers and Artificial Intelligence, Cairo University, Cairo 11562, Egypt

* Correspondence: hadeel.mostafa@guc.edu.eg

[†] Presented at the 8th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 27–30 June 2022.

Abstract: Time series forecasting is a challenging problem in the field of data mining. Deterministic forecasting has shown limitations in the field. Therefore, researchers are now more inclined towards probabilistic forecasting, which has shown a clear advantage by providing more reliable models. In this paper, we utilize seq2seq machine learning models in order to estimate prediction intervals (PIs) for a large oil production dataset. To evaluate the proposed models, Prediction Interval Coverage Probability (PICP), Prediction Interval Normalized Average Width (PINAW), and Coverage Width-based Criterion (CWC) metrics are used. Our results show that the proposed model can reliably estimate PIs for production forecasting.

Keywords: time series forecasting; prediction intervals; seq2seq; oil production



Citation: Affi, H.; Elmahdy, M.; El Saban, M.; Abu-Elkheir, M.

Probabilistic Forecasting for Oil Producing Wells Using Seq2seq Augmented Model. *Eng. Proc.* **2022**, *18*, 16. <https://doi.org/10.3390/engproc2022018016>

Academic Editors: Ignacio Rojas, Hector Pomares, Olga Valenzuela, Fernando Rojas and Luis Javier Herrera

Published: 21 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Background

A time series is a sequence of observations that is captured through time, and forecasting is the process of estimating future trends or values based on present and past values. Time series forecasting has applications in various fields, such as electricity consumption and price forecasts [1,2], wind forecasting [3], temperature forecasting [4], and several other real-life applications [5].

There are two main forecasting methods: deterministic forecasting and probabilistic forecasting [6]. Deterministic forecasting, also known as point forecasting, is the process of predicting a single deterministic value in the future, which is then compared against the target real value. However, deterministic forecasting has shown limitations in the field because no information is available about the dispersion of the actual values around the estimated values, and it is hard to tell by how much the actual value would deviate from the predicted value, which could be especially disadvantageous for complex data. Therefore, probabilistic forecasting is being explored as a forecasting method that produces potentially substantial improvement over deterministic forecasting by providing more reliable models [7]. In probabilistic forecasting, a range under which the target value should be present is predicted. This range is referred to as Prediction interval (PI).

The goal of time series analysis is to create a model that attempts to describe the behavior of the series and predict its future values. In order to facilitate the inference of information about time series, the series should be transformed to be a stationary one [8]. Moreover, most of the statistical approaches to analyzing time series data require the series to be stationary [9]. A stationary series is loosely defined as a series which has statistical properties that do not vary over time, such as mean and variance. A strict stationary series definition is too restricting; thus, a weaker version is usually used instead [10]. In order

to make the series stationary, we need to remove the trend and the seasonality. Trend represents a varying mean which can be observed in the series by the values that either keep on increasing or decreasing over time. On the other hand, seasonality is represented by a pattern which repeats itself over time, which can indicate a varying variance.

Both time series forecasting statistical and deep learning models have been discussed in the literature. Statistical models such as ARIMA are used for more precise prediction, but need experts and deep domain knowledge with vigorous analysis. On the other hand, deep learning models such as Long short-term memory (LSTM) require less knowledge in the field and less time in investigation, as there is no need to discover optimal features and parameters for the model [11]. Time series models have also adapted an attention mechanism. Attention was first introduced to solve a machine translation task [12]. The goal of attention was to overcome the shortcomings in recurrent neural networks (RNNs), as they struggled to remember long sequences. This is achieved by retaining the hidden stats at each step during decoding. Attention gives more importance to some features over the others by using weights for the features. Then, weighted sum is taken using soft max to get the sequence context for each feature.

The application of interest in this study is oil production. Oil is a traditional fossil fuels which is studied by many researchers. Even with the emergence of renewable energy such as wind, oil is still an important factor that affects the economy and plays an important role in energy investment due to the high risk, a long cost payback period and other factors accompanied by the investment in renewable energy [13].

State-of-the-art reservoir engineering forecasting techniques rely on Arps' Decline Curve Analysis (DCA) equations [14]. To estimate oil and gas reserves for current and future wells, DCA has been used as one of the prominent techniques for such a task. Arps divides the well production into two main partitions:

1. A hyperbolic curve representing the segment after an initial ramp-up period until the curve reaches a peak.
2. An exponential curve representing the decline behavior after the peak.

The curve function is summarized as follows:

$$q(t) = \begin{cases} \frac{q_i}{(1+bd_i t)^{\frac{1}{b}}} & t < t_h \\ q_h e^{-d_f t} & t \geq t_h \end{cases} \quad (1)$$

where $q(t)$ is the oil production rate in barrels/day, q_i is the initial production, d_i is the initial decline in the hyperbolic part of the equation, t is time, and b is the hyperbolic factor controlling the rate of change of the decline. After reaching a certain decline rate d_f , the curve is represented by an exponential one using q_h , the production reached by time t_h .

Decline curve modeling has been used to predict production data where the curve is fitted to the data to estimate future points. In [15], different decline models are evaluated, namely the Exponential Decline Model (SEDM) and the Logistic Growth Model (LGM), followed by the Extended Exponential Decline Model (EEDM), the Power Law Exponential Model (PLE), Doung's Model, and the Arps Hyperbolic Decline Model.

In [16], unlike the traditional trend stationarity techniques, a new method was adopted which utilizes the Arps decline curve. The trend found in the oil datasets is removed, utilizing the Arps fitted curve in an attempt to make the series stationary.

In this study, we propose a machine learning model that estimates a prediction interval for a large dataset composed of monthly oil production data of unconventional oil-producing wells. Accurate estimation of prediction intervals can play a critical role in quantifying uncertainty and to support investment and divestment decisions. The following sections are organized as follows: Section 2 discusses the experiment details and the setups used. Section 3 describes the data. Section 4 describes the evaluation metrics used to assess the prediction intervals. Section 5 includes the results and their visualization, and presents some insights. We finally conclude in Section 6.

2. Model

The machine learning model utilized in this study is a sequence-to-sequence (seq2seq) model. seq2seq is an encoder–decoder-based deep learning model. Two LSTM models are used separated by a repeat vector, which repeats the input three times. It is used as multi-step ahead forecasting, since it forecasts several steps (three) ahead in the future (future sequence). It is followed by two layers of densely connected NN of 100 units and 1 unit applied using TimeDistributed layers.

Using the model, we test different setups, which include trend removal and attention mechanism. Quantile loss is utilized to create the upper and lower bounds of the PIs, as well as the 0.5 quantile (p50).

Trend removal is used to make the series stationary by utilizing the special trend accompanied by oil production series. First, the sequence is fitted using a hyperbolic-to-exponential Arps decline curve. Then, trend removal is simply achieved by taking the difference between the original series with the Arps fitted curve, as shown in Figure 1. Regarding the attention mechanism, we implemented a simple attention layer using keras [17] following the attention mechanism introduced in [12].

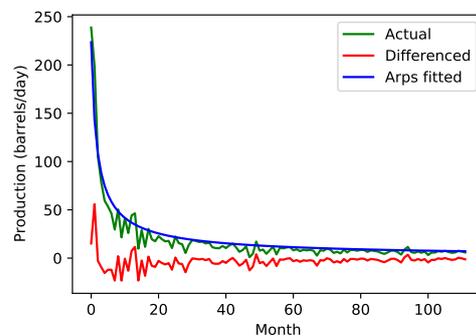


Figure 1. Trend removal by taking the difference between the actual curve and the Arps fitted curve.

3. Data

Experiments are conducted on a dataset consisting of sequences of production data obtained over successive months for producing wells from all US oil and gas basins. These sequences are represented by the number of oil barrels per day from oil horizontal wells. Only the data post the peak production are used, bearing in mind that typically, the data prior to the peak production in the reservoir engineering domain are studied independently of the rest of the data. The total number of wells in our experiment is 60,000, where 50,000 were used to train the model and 10,000 are withheld for the testing phase. The sliding window technique is leveraged to cover all months and make the input sequence consistent in size where sequences of nine consecutive months are taken; six months are used as features, and three as targets. Accordingly, the training set consists of 1,596,240 sequences and the test set consists of 46,386 sequences. We aim to estimate an interval by employing the quantile loss using the 0.5 quantile for the lower bound and the 0.95 quantile for the upper one to achieve 90% of the prediction interval.

4. Evaluation Metrics

Different metrics are utilized to evaluate the predicted prediction intervals. The most commonly used metrics are: Prediction Interval Coverage Probability (PICP) and Prediction Interval Normalized Average Width (PINAW) [2,18]. PICP is a method that measures the probability of a specific target existing within the predicted interval. It is defined as follows:

$$PICP = \frac{1}{N} \sum_{i=1}^N \epsilon_i \quad (2)$$

where ϵ_i is defined as

$$\epsilon_i = \begin{cases} 1, & \text{if } x_i \in [L_i, U_i] \\ 0, & \text{if } x_i \notin [L_i, U_i] \end{cases}$$

L_i and U_i represent the lower and upper bound of the prediction interval, respectively; N is the number of samples, and x_i is the target.

The results are enhanced significantly by increasing the PICP value. However, the width of the PI has an impact on the prediction. Increasing the PI width by moving the lower and upper bounds further apart to include more targets negatively affects the prediction significance, as the decision-makers will have little information to base their decision upon. PINAW, also known in the literature as Normalized Mean Prediction Interval Width (NMPiW), was introduced to overcome this flaw; it is a metric that measures the width of the interval. It is commonly used in the literature with the investigation of probabilistic forecasting. PINAW is the average width of the several predicted PIs normalized by the width of the target, and it is defined as follows:

$$PINAW = \frac{1}{N * R} \sum_{i=1}^N (U_i - L_i) \quad (3)$$

where R is the range of the target; in other words, it is the maximum minus the minimum target.

Hence, the smaller the value of the PINAW, the better the results are.

It is desirable to have a narrow PI, which can be obtained by targeting a narrower interval while considering the quantiles. However, that has conflicted interests with having a large number of target points, which can be obtained by making the PI wider. Therefore, a Coverage Width-based Criterion (CWC) is introduced in the literature [19]. CWC is defined as

$$CWC = PINAW * (1 + \gamma(PICP) * e^{-\eta(PICP - \mu)}) \quad (4)$$

where $\gamma(PICP)$ is defined as follows:

$$\gamma(PICP) = \begin{cases} 1, & \text{if } PICP < \mu \\ 0, & \text{if } PICP \geq \mu \end{cases}$$

While the hyper-parameter μ is the target PICP value and η is the penalty for having a PICP value less than the target.

The value of η should be large to provide a high penalty for non-sufficiently-informative PIs. Having reached the target PICP, the CWC will have the same value as the PINAW and, in that case, it is safe to assume an informative PI. On the other hand, a smaller PICP will lead to high CWC values caused by the penalty η in the exponential term in equation 4. Hence, a small CWC is targeted.

5. Results and Discussion

The results are shown in Table 1. Adding attention only and 90% PI by choosing 0.05 and 0.95 quantiles, we are able to obtain 90.5% PICP, 9.4% for PINAW and, 0.094 CWC when the expected PICP is set to 90% and the penalty parameter is equal to 50 [20]. On the other hand, by applying trend removal only, we obtain 85.9% PICP, 6.9% PINAW, and 0.605 CWC. Additionally, when using both trend removal and attention, we get 85.4% PICP, 6.7% PINAW and 0.729 CWC. The CWC value increases with non-satisfactory PICP, which means when PICP is less than the expected value, the CWC will be significantly greater than PINAW. CWC will be equal to PINAW when the PICP value is greater than or equal to the expected value. Thus, we can deduce that the smaller the CWC value is, the better the prediction is. The very slight improvement upon using the attention can be attributed to the sequence under investigation; the sequence is not long enough to emphasize the enhancement.

Table 1. Seq2seq results for different setups in percentage using PICP, PINAW and CWC evaluation metrics. Arps in the set-up means trend is removed using Arps differencing. The μ is set to 0.9 for 90% PI and 0.8 for 80% PI.

Set-Up	PICP	PINAW	CWC $\eta = 50$
90 PI + Arps	85.9	6.9	60.5
90 PI + Attention	90.49	9.4	9.4
90 PI + Arps + Attention	85.42	6.7	72.9
90 PI	90.93	9.8	9.8
80 PI	82.16	6.4	6.4

The result in our work regarding using Arps differencing confirms the results in [16], which can be found in Table 2. From our results in Table 1, it is clear that using Arps differencing yields a narrower width compared to keeping the trend. We can also see that using Arps differencing is better than choosing a narrower PI in regards to PICP, which represents the coverage probability of the PI when both yield a narrower width, as shown in Figure 2.

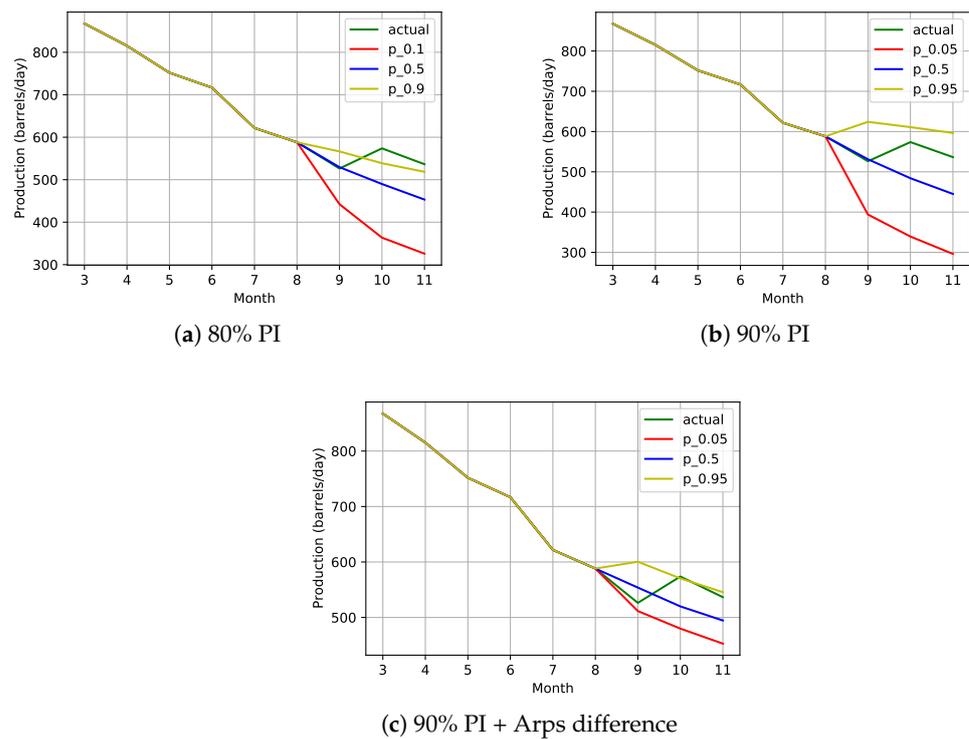


Figure 2. Arps differencing yields a narrower width compared to keeping the trend. It also has more PI coverage probability compared to a narrower width.

Table 2. Gradient Boosting for regression results in percentage for different setups using PICP and PINAW evaluation metrics in [16]. Each step represents a predicted month.

Set-Up	Step 1		Step 2		Step 3		Aggregation	
	PICP	PINAW	PICP	PINAW	PICP	PINAW	PICP	PINAW
Direct 80	82.4	4.7	81.4	6.6	81.0	8.0	81.64	5.2
Direct 90	91.1	7.9	90.7	11.0	90.6	12.9	90.80	8.6
Arps Difference Direct 90	86.6	6.1	85.1	8.3	84.1	9.7	85.23	6.5

6. Conclusions and Future Work

The proposed machine learning model uses monthly production data as features. The sliding window technique is utilized to cover all production months. The model applies the Arps differencing technique to reduce the variation in the statistical properties that make the series non-stationary. Moreover, we investigate the effect of applying the attention mechanism to the model. The seq2seq deep learning model is utilized by taking the first n production points starting from the peak production to predict a $p50$ production curve, as well as a prediction interval in which the next m production points observed will most likely lie. Prediction intervals are estimated besides the $p50$ in order to capture uncertainties lacking in the deterministic forecasts. Further investigation applying different architectures and machine learning models and observing the outcome after removing other special trends that may accompany other datasets is warranted.

Author Contributions: Conceptualization, H.A., M.E., M.E.S. and M.A.-E.; methodology, H.A., M.E. and M.E.S.; software, H.A.; validation, H.A.; resources, H.A., M.E., M.E.S.; writing original draft preparation, H.A.; writing—review and editing, M.E., M.E.S. and M.A.-E.; visualization, H.A.; supervision, M.E., M.E.S. and M.A.-E.; project administration, M.E., M.E.S. and M.A.-E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Production data are publicly available but a third party service (Conduit Resources) was used to collect the scattered data from many sources.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zeng, Y.-R.; Zeng, Y.; Choi, B.; Wang, L. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. *Energy* **2017**, *127*, 381–396. [[CrossRef](#)]
2. Khosravi, A.; Nahavandi, S.; Creighton, D. Quantifying uncertainties of neural network-based electricity price forecasts. *Appl. Energy* **2013**, *112*, 120–129. [[CrossRef](#)]
3. Soman, S.S.; Zareipour, H.; Malik, O.; Mandal, P. A review of wind power and wind speed forecasting methods with different time horizons. In Proceedings of the North American Power Symposium 2010, Arlington, TX, USA, 26–28 September 2010; pp. 1–8.
4. Romeu, P.; Zamora-martínez, F.; Botella-Rocamora, P.; Pardo, J. Time-series forecasting of indoor temperature using pre-trained deep neural networks. In *Proceedings of the International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 451–458.
5. Gooijer, J.G.D.; Hyndman, R.J. 25 years of time series forecasting. *Int. J. Forecast.* **2006**, *22*, 443–473. [[CrossRef](#)]
6. Wang, H.; Yi, H.; Peng, J.; Wang, G.; Liu, Y.; Jiang, H.; Liu, W. Deterministic and probabilistic forecasting of photovoltaic power base don deep convolutional neural network. *Energy Convers. Manag.* **2017**, *153*, 409–422. [[CrossRef](#)]
7. Camporeale, E.; Chu, X.; Agapitov, O.; Bortnik, J. On the generation of probabilistic forecasts from deterministic models. *Space Weather* **2019**, *17*, 455–475. [[CrossRef](#)]
8. Kirchgässner, G.; Wolters, J.; Hassler, U. *Introduction to Modern Time Series Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
9. Chang, F.; Huang, H.; Chan, A.H.; Man, S.S.; Gong, Y.; Zhou, H. Capturing long-memory properties in road fatality rate series by an autoregressive fractionally integrated moving average model with generalized autoregressive conditional heteroscedasticity: A case study of florida, the united states, 1975–2018. *J. Saf. Res.* **2022**, *81*, 216–224. [[CrossRef](#)] [[PubMed](#)]
10. Nason, G.P. Stationary and non-stationary time series. In *Statistics in Volcanology*; Geological Society of London: London, UK, 2006; Volume 60.
11. Cecaj, A.; Lippi, M.; Mamei, M.; Zambonelli, F. Comparing deep learning and statistical methods in forecasting crowd distribution from aggregated mobile phone data. *Appl. Sci.* **2020**, *10*, 6580. [[CrossRef](#)]
12. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
13. Zhang, Y.-J.; Chen, M.-Y. Evaluating the dynamic performance of energy portfolios: Empirical evidence from the dea directional distance function. *Eur. J. Oper. Res.* **2018**, *269*, 64–78. [[CrossRef](#)]
14. Ma, X.; Liu, Z. Predicting the oil production using the novel multivariate nonlinear model based on arps decline model and kernel method. *Neural Comput. Appl.* **2018**, *29*, 579–591. [[CrossRef](#)]
15. Manda, P.; Nkazi, D.B. The evaluation and sensitivity of decline curve modelling. *Energies* **2020**, *13*, 2765. [[CrossRef](#)]

16. Afifi, H.; Elmahdy, M.; Saban, M.E.; Abu-Elkheir, M. Probabilistic time series forecasting for unconventional oil and gas producing wells. In Proceedings of the 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 24–26 October 2020; pp. 450–455.
17. The Base Layer Class. Available online: <https://keras.io/api/layers/baselayer/layer-class> (accessed on 30 May 2021).
18. der Meer, D.W.V.; Widén, J.; Munkhammar, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1484–1512. [[CrossRef](#)]
19. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Trans. Neural Netw.* **2010**, *22*, 337–346. [[CrossRef](#)] [[PubMed](#)]
20. Petneházi, G. Recurrent neural networks for time series forecasting. *arXiv* **2019**, arXiv:1901.00069.