2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

# Oil Price Forecasting Using Supervised GANs with Continuous Wavelet Transform Features

Zhaojie Luo*, Jinhui Chen*, Xiao Jing Cai†, Katsuyuki Tanaka†,
Tetsuya Takiguchi*, Takuji Kinkyo†, Shigeyuki Hamori†
*Graduate School of System Informatics, Kobe University
Email: {luozhaojie, ianchen}@me.cs.scitec.kobe-u.ac.jp
†Graduate School of Economics, Kobe University

***Abstract*—This paper proposes a novel approach based on a supervised Generative Adversarial Networks (GANs) model that forecasts the crude oil prices with Adaptive Scales Continuous Wavelet Transform (AS-CWT). In our study, we first confirmed that the possibility of using Continuous Wavelet Transform (CWT) to decompose an oil price series into various components, such as the sequence of days, weeks, months and years, so that the decomposed new time series can be used as inputs for a deep-learning (DL) training model. Second, we find that applying the proposed adaptive scales in the CWT method can strengthen the dependence of inputs and provide more useful information, which can improve the forecasting performance. Finally, we use the supervised GANs model as a training model, which can provide more accurate forecasts than those of the naive forecast (NF) model and other nonlinear models, such as Neural Networks (NNs), and Deep Belief Networks (DBNs) when dealing with a limited amount of oil prices data.**

## I. Introduction

Oil price forecasting has many implications for the economic growth of countries as well as providing useful information that helps international investors to diversify risk. According to BP's Statistical Energy Outlook, crude oil is a vital fuel, accounting for 32.9% of global energy consumption in 2016, and will continue to play an important role until 2035. It is generally accepted that the oil price fluctuations have a significant influence on macroeconomic aggregates, such as the GDP and inflation of oil-exporting and -importing countries, as one of the most actively traded commodities in the world [1]. Thus, it is important to focus on improving the forecasting accuracy of oil prices for both real economy and financial markets. However, oil price forecasting is rather challenging because the crude oil prices are usually considered to be a nonlinear and non-stationary time series, and are interactively affected by many factors.

Research on crude oil price forecasting has lasted for decades, with many machine learning techniques being utilized to mine the inner complexity of oil prices. Among these approaches, neural networks (NNs) have been commonly used because NN models can create a breakthrough opportunity in the analysis of the non-linear behavior of the crude oil prices [2], [3]. For example, Moshiri *et al.* [4] compared linear economic models (ARMA and GARCH) with nonlinear NN models, and found that NNs are superior and produce a more statistically significant forecasting. Wang *et al.* [5] forecast monthly prices by using an NNs-based model, and claimed superior performance by the model. However, shallow architecture models, such as the the NN-based forecasting models mentioned above, cannot model the complex patterns and volatile behaviors of oil prices, which are influenced by numerous factors (Bengio *et al.* [6]).

Recently, the deep-learning (DL) approach is becoming a mainstream of machine learning technique, and has dramatically improved the performance of various nonlinear modeling tasks due to the multi-layers architecture. Hinton *et al.* [7] proposed a greedy layer-wise training strategy which solves the training problem in deep neural networks (DNNs). And, Yu *et al.* [8] have applied the DL approach to the oil price forecasting. However, there are two problems for oil prices forecasting when using the DL models. First, deep learning models are restricted to problems with moderate dimensions for training data. The original oil price is a one-dimensional sequence that is not suitable for DL approaches. Thus, the oil price data need to be transformed to high dimensional data before training the DL models. Second, DL models need sufficient data for training. But, compared with the size of training data used in speech signal processing or image processing tasks, the oil price data is insufficient for DL models.

In this paper, to overcome the two problems mentioned above, we propose a novel method that uses adaptive scales CWT (AS-CWT) to decompose one dimensional oil price data to high-dimensional features, and then train them with the supervised GANs model.

Wavelet analysis has recently been used in the economic fields of time-series analysis, such as business cycle synchronization, commodities, and to study the co-movement among financial markets. In oil price forecasting, Jammazi *et al.* [9] combined the wavelet transform and a NN to forecast the crude oil monthly price. Tang *et al.* [10] constructed a multiple-wavelet recurrent NN model to analyze crude oil monthly prices. Different from these papers, which predict the monthly oil prices, in our paper, we aim to forecast the daily oil price series by using the AS-CWT method with supervised GANs models. The proposed AS-CWT method can systematically capture the oil prices of different temporal scales by using adaptive scales, which can then represent different oil prices levels ranging from daily prices to yearly prices levels, but better optimized.

Moreover, to overcome the difficulty of a limited amount of training data, we propose a supervised GANs model. The GANs are able to take advantage of an adversarial loss forcing the generated data to be indistinguishable from real data. This is particularly powerful for image-generation tasks, but it has not yet begun to be applied in economic fields. Thus, due to supervised learning's ability to regularize the training process of a GANs model, we developed a novel supervised GANs model that enables the price-forecasting function to be trained from two sets of labeled oil prices from recent to future domains.

In the remaining sections of this paper, previous literature, including CWT and GANs, are reviewed in Section II. Then, we describe our proposed oil price-forecasting method in Section III. Section IV gives the detailed stages process of experimental evaluations, and Section V presents our conclusions.

## II. Related works

Our oil price-forecasting system uses GANs combined with supervised learning to capture high-order conversion-friendly CWT oil prices features. In this section we briefly review how to decompose one dimensional oil prices to high-dimensional CWT oil features using continuous wavelet transform, and introduce the related fundamental GANs model.

### *A. Continuous Wavelet Transform*

It is well known that oil prices forecasting is influenced both by, short-term dependencies, such as daily levels, and by long-term dependencies, such as yearly levels. A CWT is used to decompose a signal into wavelets and is an excellent tool for mapping the changing properties of non-stationary signals. Consequently, CWT is the best method for the analysis of international crude oil. The continuous wavelet transform of the oil prices is defined by

$$W(\tau, t) = \tau^{-1/2} \int_{-\infty}^{\infty} p(x)\, \psi\left(\frac{x-t}{\tau}\right) dx \tag{1}$$

$$\psi(t) = \frac{2}{\sqrt{3}} \pi^{-1/4} \left(1 - t^2\right) e^{-t^2/2}, \tag{2}$$

where $\tau$ is the scaling factor, $t$ is the translating factor, $p(x)$ represents the input recent oil price series and $\psi$ is the Mexican hat mother wavelet. The original signal $p$ can be recovered from the wavelet representation $W(p)$ by inverse transform [11]:

$$p(t) = \int_{-\infty}^{\infty} \int_{0}^{\infty} W(p)(\tau, x) \tau^{-5/2} \psi(\frac{t-x}{\tau}) dx d\tau \tag{3}$$

The coefficients of the continuous wavelet transform have a significant amount of redundant information. Therefore, it is reasonable to sample the coefficients in order to reduce redundancy. Thus, in our proposed AS-CWT method, we decompose an oil price series into useful sets (daily level, weekly level, monthly level and yearly level) of components results using the adaptive scales.

### *B. Generative Adversarial Networks*

Generative Adversarial Networks (GANs) [12], [13] have achieved impressive results in image generation [12], [14], image editing [13], and representation learning [15]. Key to the success of the GANs is learning a generator distribution $P_{G(x)}$ that matches the true data distribution. It consists of two networks: a generator $G$ that transforms noise variables $z \sim P_{Noise(z)}$ to data space $x = G(z)$ and a discriminator $D$ that assigns probability $p = D(x)$ when $x$ is a sample from the $P_{Data(x)}$ and assigns probability $1-p$ when $x$ is a sample from the $P_{G(x)}$. In a GAN, $D$ and $G$ play the following two-player minimax game with the value function $V(G, D)$:

$$\begin{aligned} \min_G \max_D V(D, G) = &\, E_{x \sim p_{data}(x)}[\log D(x)] \\ &+ E_{x \sim p_z(z)}[\log(1 - D(G(z)))] \end{aligned} \tag{4}$$

This enables discriminator $D$ to find the binary classifier that provides the best possible discrimination between true and generated data, and simultaneously enables generator $G$ to fit $P_{Data(x)}$. Both $G$ and $D$ can be trained using back-propagation. Although the GANs models have great effectiveness for image-generation tasks, unlike normal object image generation, the dissimilarities between the source (recent oil prices series) and target (future oil prices series) are very small. Therefore, without supervised learning, the source features sometime are hard to be regularized to the target features due to the problem of insufficient data. Thus, in our forecasting system, we propose GANs combined with supervised learning.

## III. Oil price forecasting using supervised GANs with CWT features

### *A. Adaptive Scales CWT*

In the current paper, we apply an adaptive scales method to decompose an oil price series using a wavelet transform. As shown in the left part of Fig. 1, there are two main steps in calculating the adaptive scales. 1) We investigate the variability in each temporal level as a rich source of information for studying the degree of impact of every level in oil prices forecasting as a function of $influencing\ strength$, and, 2) calculate adaptive scaling factors with the $influencing\ strength$. The details of the steps in this process are described below.

1) Let $x^* \in \{X_w, X_m, X_y\}$ be values of each the temporal level with $X_w$, $X_m$ and $X_y$ representing the week, month and year levels, respectively. Because the oil price will not change over the weekends and on holidays, we denote week level $X_w$=5, month level $X_m$=20 and year level $X_y$=240, respectively. Then, we calculate each temporal level's $influencing\ strength$ which represents the proportion of influence among all the temporal levels. As shown in the first part of Fig. 1, we define some functions: relative distance ($RD$), different area ($Da$) and basic area ($Ba$). These functions are used for calculating the $influencing\ strength$. For example, $Da(X_w)$ and $Ba(X_w)$ represent the different area and basic area of the week level between the recent oil price and future oil price.

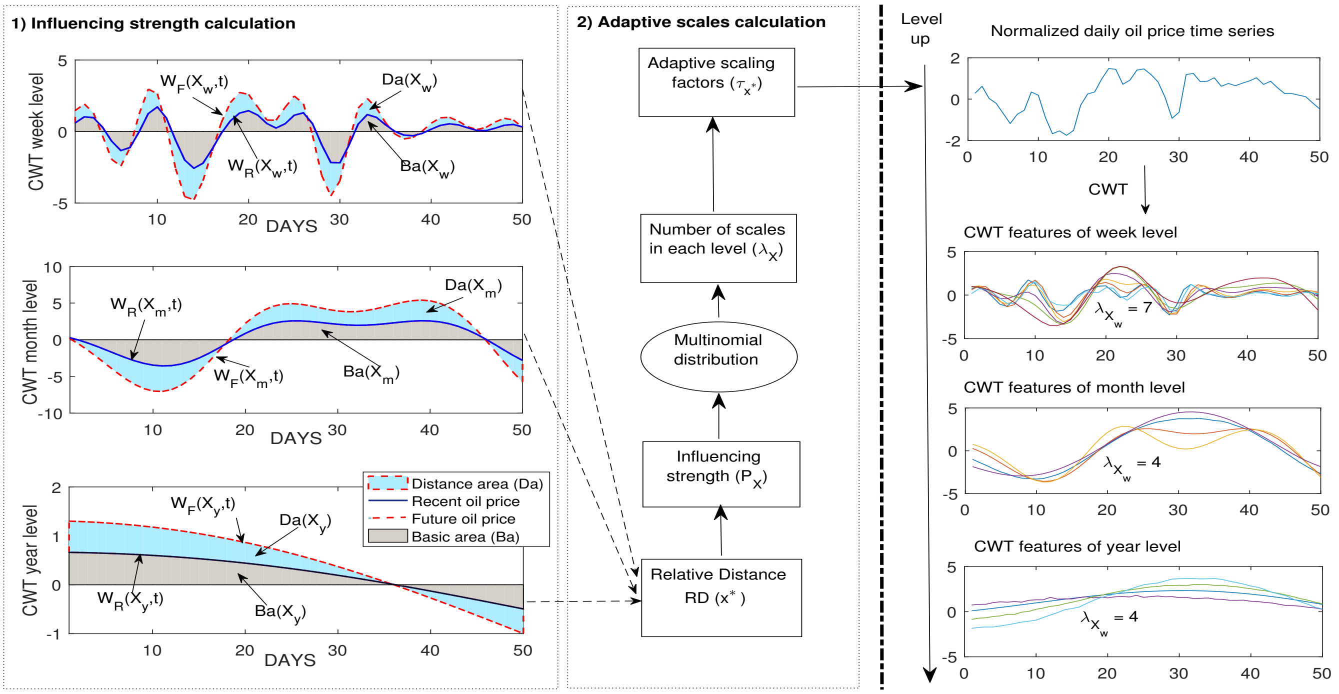


Fig. 1. Illustration of calculating the adaptive scales for CWT and using them to decompose the oil prices series. On the left, the two main steps in calculating the adaptive scaling factors are shown, and on the right, samples of CWT features decomposed by AS-CWT are shown.

Now, let's consider how the relative distance function $RD$ calculates the relative distance between the recent oil price and future oil price in each temporal level such that:

$$RD(x^*) = \frac{Da(x^*)}{Ba(x^*)} \tag{5}$$

where $Da$ and $Ba$ formulas are calculated as below:

$$Da(x^*) = |W_R(x^*) - W_F(x^*)| \tag{6}$$

$$Ba(x^*) = \min\left(|W_R(x^*)|, |W_F(x^*)|\right) \tag{7}$$

In the $Da$ and $Ba$ functions, $W_R(x^*)$ represents the continuous wavelet transform function of a recent oil price series in different temporal level $x^*$, and $W_F(x^*)$ uses the future oil price series as input. Their transform functions are defined by

$$\begin{aligned} W_R(x^*,t) &= (x^*)^{-1/2} \int_{-\infty}^{\infty} p_r \psi\left(\frac{x-t}{x^*}\right) dx \\ W_F(x^*,t) &= (x^*)^{-1/2} \int_{-\infty}^{\infty} p_f \psi\left(\frac{x-t}{x^*}\right) dx \end{aligned} \tag{8}$$

$$\psi(t) = \frac{2}{\sqrt{3}} \pi^{-1/4} \left(1 - t^2\right) e^{-t^2/2}, \tag{9}$$

where $\psi$ is the Mexican hat wavelet, $p_r$ and $p_f$ represent the recent and future oil prices series input signal, respectively. And, the $influencing\ strength$ of each temporal level can be ranked by

$$P_{x^*} = \frac{RD(x^*)}{\sum_{x^* \in X} RD(x^*)} \tag{10}$$

Then, we can draw the optimized number of scales for CWT in each temporal level with the $influencing\ strength$ from a multinomial distribution:

$$\begin{aligned} \lambda_{\mathbf{X}} &\sim Multinomial(N, \mathbf{P_X}), \\ \lambda_{x^*} \in \lambda_{\mathbf{X}} &= (\lambda_{X_w}, \lambda_{X_m}, \lambda_{X_y}), \\ P_{x^*} \in \mathbf{P_X} &= (P_{X_w}, P_{X_m}, P_{X_y}) \end{aligned} \tag{11}$$

where $N$ is the total number of scales, which can be set in different values, vectors $\mathbf{P_X}$ are made up of all the $influencing\ strengths$, and $\lambda_{\mathbf{X}}$ represents the aggregation of the number of scales in all temporal levels. Therefore, $\lambda_{x^*}$ can represent the number of scales in each temporal level.

2) The second step is applying the number of scales of each temporal level calculated in first step to the CWT function. As we know, weeks are made up of days, months are made up of weeks and years are made up of months. Thus, we can use the $\lambda_{x^*}$ (number of scales) to calculate the adaptive scaling factor of each temporal level as:

$$\begin{aligned} \tau_{x^*} &= (\tau_{X_w}, \tau_{X_m}, \tau_{X_y}), \\ \tau_{X_w} &= X_w + X_d * i_w,\ i_w = 1, ..., \lambda_{X_w}, \\ \tau_{X_m} &= X_m + X_w * i_m,\ i_m = 1, ..., \lambda_{X_m}, \\ \tau_{X_y} &= X_y + X_m * i_y,\ i_y = 1, ..., \lambda_{X_y} \end{aligned} \tag{12}$$

where $\tau_{X_w}$, $\tau_{X_m}$ and $\tau_{X_y}$ represent the adaptive scaling factor for week, month and year, respectively. They are all calculated by the number of scales $\lambda_{x^*}$ and the values of the previous level. For example, scaling factors for yearly level ($\tau_{X_y}$) are calculated by the number of scales of years ($\lambda_{X_y}$) and monthly values $X_m$. In the first step, we have defined $X_w$, $X_m$ and $X_y$ to 5, 20 and 200, respectively which can represent the basic value of each temporal level. Here, $X_d = 1$ represents the value of daily level.

After calculating the scaling factors of each temporal level, we adopt CWT to decompose the contour oil price series with these scaling factors and our oil prices series can be represented by separating components given by

$$W(\tau_{x^*}, t) = (\tau_{x^*})^{-1/2} \int_{-\infty}^{\infty} p(x)\, \psi \left( \frac{x-t}{\tau_{x^*}} \right) dx \tag{13}$$

The original signal is approximately recovered by

$$p(t) = \int_{-\infty}^{\infty} \int_{0}^{\infty} W(p)(\tau_{x^*}, x) \tau^{-5/2} \psi(\frac{t-x}{\tau_{x^*}}) dx d\tau \tag{14}$$

### B. Training model

Before training our proposed supervised GANs model, we reshape the AS-CWT features to 2D features of $N \times N$ size. $N$ is the total number of scales, which is set in the AS-CWT features process part (Sec. III-A). As shown in Fig. 2, $N$ is set to 32. The supervised GANs is comprised of two generators ($G_z$, $G_x$) and a discriminator ($D_y$). Here, $G_z$ is responsible for generating realistic samples close to the content of the target dataset $y \sim P_{Data(y)}$ from the noise features $z \sim P_{Noise(z)}$, while $G_x$ is responsible for generating realistic samples from the input dataset $x \sim P_{Data(x)}$. $D$ is responsible for determining true and generated data to use in discrimination. In this setting, the objective function is written as

$$\begin{aligned} L_{GAN} &= L_{G_x}(G_x, D_y, x, y) + L_{G_z}(G_z, D_y, z, y), \\ L_{G_x} &= E_{y \sim P_{data(y)}}[\log D(y)] \\ &\quad + E_{x \sim P_{Data(x)}}[\log(1 - D(G_x(x)))] \\ L_{G_z} &= E_{y \sim P_{data(y)}}[\log D(y)] \\ &\quad + E_{z \sim P_{Noise(x)}}[\log(1 - D(G_z(z)))] \end{aligned} \tag{15}$$

where $G_z$ and $G_x$ try to minimize this objective against an adversarial $D_y$ that tries to maximize it, and our final objective is

$$(G_z)^*, (G_x)^* = arg \max_{D_y} \min_{G_z, G_x} L_{GAN}(G_z, G_x, D_y) \tag{16}$$

Without $x$, the GANs model could still learn a mapping from $z$ to $y$, but would produce nondeterministic outputs, and, therefore, fail to regularize the training process due to the problem of insufficient data. Thus, in this model, adding the supervised process can enhance robustness of the GANs' training.

The training procedure of the proposed approaches is briefly portrayed in Alg. 1. Throughout the training process, generators, $G_x$ and $G_z$ are optimized to learn the generated oil price which cannot be distinguished from target future oil price by corresponding discriminators $D_y$.

**Algorithm 1** Training procedure of supervised GANs

**Require:** AS-CWT features ($32 \times 32$) sets processed from source recent oil price $x$ and target future oil price $y$, generator $G_x$ with generator parameters $\theta_x$ and discriminator parameters $w_x$, generator $G_z$ with generator parameters $\theta_z$, and discriminator parameters $w_z$, batch size $m$, and the epochs $n$.

Initialize the parameters $\theta_x$, $\theta_z$, $w_x$ and $w_z$, randomly.
**repeat**
  **for**($i = 1; i < n+1; i = i+1$) **do**
    sample AS-CWT features $z_k \subseteq z$, $x_k \subseteq x$, $y_k \subseteq y$, $k \in \{1, ..., m\}$
    update $w_x$, $\theta_x$ to minimize $\frac{1}{m} \sum_{k=1}^{m} L_{G_x}(x_k, y_k)$
    update $w_z$, $\theta_z$ to minimize $\frac{1}{m} \sum_{k=1}^{m} L_G(z_k, y_k)$
  **end for**
  update $w_x$, $w_z$ $\theta_x$, $\theta_z$ to minimize $\frac{1}{m} \sum_{k=1}^{m} L_{GAN}$
**until** convergence

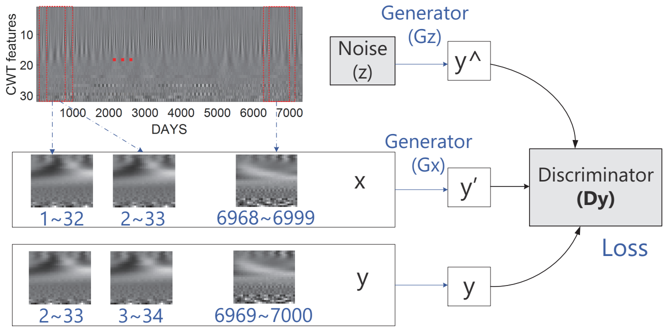


Fig. 2. Illustration of feature extraction and calculating the loss of supervised-GANs. $x$ and $y$ represents the 2D-features processed from recent oil price and future oil price, respectively. $y^\wedge$ and $y'$ represent the generated features from noise features and the CWT features of recent oil prices, respectively.

## IV. Experiments

### A. Experimental Setup

In this study, we use Brent crude oil future of the front-month futures contracts (ICE futures Europe). The data set covers the period from June 27, 1988, to November 4, 2016, consisting of 7300 observations. In the datasets, 7000 days' oil price values were chosen as training data, and the remaining 300 days' values were chosen for the evaluation.

To evaluate the proposed supervised GANs model, we compared the results with several state-of-the-art models, which are listed below.

- **Naive Forecast (NF):** The traditional economic models are based on the navie forecast. As described in [16], in the NF, the oil price tomorrow is equal to the oil price today. And, at the same time, the probability of an increase (or decrease) in the oil price the next day is just 50%.
- **NNs:** Nonlinear neural network with shallow architecture [17].
- **DBNs:** Deep neural networks with multi-layers architecture; among the deep-learning models, Deep Belief Networks (DBNs) have demonstrated excellent performance [18].

- **Supervised-GANs (the proposed method):** This is our proposed method, which uses GANs to train oil price features with the supervised learning.

To evaluate the effectiveness of the proposed AS-CWT method, we compared the results when using the AS-CWT method and when not using the AS-CWT method for all non-linear models (NNs, DBNs, supervised GANs).

### *B. Training Procedure*

Table I details the network architectures of the generator $G_x$, $G_z$, and discriminator $D_y$. The symbols ↓ and ↑ indicate down-sampling and up-sampling, respectively. To upscale and downscale, we used convolutions and backward convolutions with stride 1, respectively.

In the generator networks ($G_x$, $G_z$), similar to Johnson *et al.* [19], we use batch normalization (BNorm) [20] and all convolutional layers are followed by ReLU nonlinearities [21] with the exception of the output layer. Input and output are set as the 2D training features ($32 \times 32$). Each generator network contains two stride-1 convolutions to downsample the input, followed by two residual blocks [22], and two fractional stride convolutions with stride $\frac{1}{2}$ to upsample.

For the discriminator network ($D_y$), we use a convolutional PatchGAN classifier [23]. The patch size at which the discriminator operates is fixed at $10 \times 10$.

TABLE I
DETAILS OF NETWORK ARCHITECTURES OF $F$, $G$, $D_x$, $D_y$.

| |
|---|
| $G_z$ (Input: $32 \times 32$ features with random noise, Output: $32 \times 32$ generated features |
| 2×2 32 conv.↓, BNorm, ReLU |
| 2×2 16 conv.↓, ReLU |
| *residual blocks* $\begin{bmatrix} 2\times 2 & 16 & conv. & ReLU \\ 2\times 2 & 16 & conv. & ReLU \end{bmatrix} \times 2$ |
| 2×2 16 conv.↑, ReLU |
| 2×2 32 conv.↑, BNorm, ReLU |
| $G_x$ (Input: $32 \times 32$ AS-CWT features processed by recent oil price $x$, Output: generated $32 \times 32$ AS-CWT features) |
| 2×2 32 conv.↓, BNorm, ReLU |
| 2×2 16 conv.↓, ReLU |
| 2×2 16 conv.↑, ReLU |
| 2×2 32 conv.↑, BNorm, ReLU |
| $D_y$ (Input: $32 \times 32$ AS-CWT features, Output: 1 Probability) |
| 2×2 16 conv.↓, ReLU |
| 2×2 32 conv.↓, BNorm, ReLU |
| 128 fully connected, BNorm, ReLU |
| 1 fully connected, sigmoid |

During preprocessing, we normalized the recent oil price series and future oil price series to zero-mean and unit-variance. Then, we transformed them to $32 \times 32$ AS-CWT features. When training $G_x$, $G_z$ and $D_y$, we use the Adam optimizer [24] with a mini-batch size. The learning rate was set to 0.0001 for $G_x$, 0.0002 for $G_z$, and 0.0001 for $D_y$, respectively. The momentum term was set to 0.5.

To clarify the characteristics of our proposed method, as described above, we implemented NNs and DBNs model for comparison. The NNs model has 2 hidden layers, and the numbers of units in the input, hidden, and output layers are [N, 2*N, 2*N, N]. We use the DBNs model proposed by Nakashika *et al.* [25], which contains two different DBNs for source oil price (unit [N, 2*N, N]) and target oil price ([N, 2*N, N]), and the connected NNs (unit [N, 2*N, N]). Here, $N$ is the dimensions of the input and output features.

TABLE II
DA AND RMSE RESULTS FOR MODEL COMPARISON.

| | NF | NNs | DBNs | GANs | NNs+ | DBNs+ | GANs+ |
|---|---|---|---|---|---|---|---|
| RMSE | 0.914 | 0.873 | 0.816 | 0.764 | 0.854 | 0.772 | 0.768 |
| DA | 0.414 | 0.455 | 0.515 | 0.466 | 0.475 | 0.523 | 0.572 |

Notes: NF denotes Naive Forecast; NNs, DBNs and S-GAN represent the NNs model, DBNs model and supervised GANs model without AS-CWT method; (+) represents using the AS-CWT method.

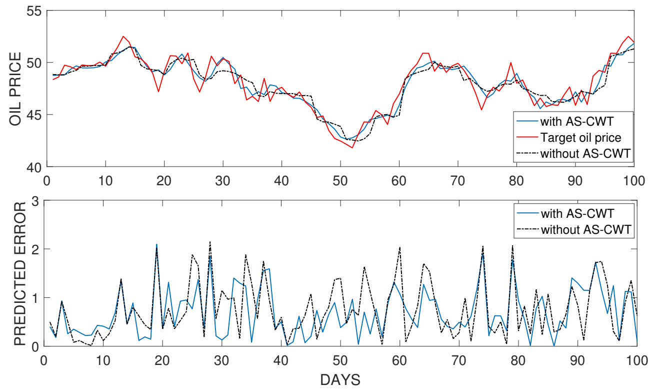


Fig. 3. Top: Target oil price (red) and the predicted price by supervised GANs without AS-CWT (black) and with AS-CWT (blue), Bottom: The predicted errors

### *C. Results*

To evaluate the forecasting performance, we calculate the root mean square error (RMSE) and directional accuracy (DA) between the actual values and predicted values, which are often used in the literature [9], [26]. The RMSE can reflect the disparity between the actual values and predicted values, while the DA can represent the directional accuracy of each day between the actual data and predicted data, which can be expressed as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(V_t^a - V_t^p)^2} \tag{17}$$

$$DA = \frac{1}{N}\sum_{t=1}^{N} Z_t, \quad Z_t = \begin{cases} 1 & (V_t^a - V_{t-1}^a)(V_t^p - V_{t-1}^p) \geq 0 \\ 0 & otherwise \end{cases} \tag{18}$$

where $V_t^a$ and $V_t^p$ denote the actual value and predicted value, respectively. $N$ represents the number of days in the testing data. A lower RMSE means a smaller difference between the actual value and predicted value, while a lager DA represents a higher directional accuracy of the predicted value. Thus, the lower RMSE and higher value of DA represent a better forecasting performance of the model.

Table II shows the results of each model for oil price forecasting. From Table II, 1) we can see that all the NNs models achieve larger DA and smaller RMSE values than the NF model, confirming that the AI-based forecasting model can provide greater efficiency and higher accuracy. 2) Comparing the results of NNs and DBNs, we find that implying the model with deep layers provides higher forecasting accuracy than the shallow architecture model. The result is in line with [6] Bengio (2009). 3) When comparing the results of supervised GANs with the conventional methods, the proposed supervised GANs can obtain a better RMSE result, which represents a better oil value forecasting effect, but it obtains worse results than a DBNs model for DA when not using AS-CWT features. We recognize that using original oil prices can mitigate the over-smoothing problem, but sometimes it is hard to regular training process of GANs. 4) When comparing the results of models with using and without using AS-CWT features, we can find that using the AS-CWT features can improve the effectiveness of all models, especially supervised GANs in regard to DA. This indicates that the AS-CWT features make up for the shortage of convergence and stability of GANs, and improve the directional accuracy of oil price forecasting.

Fig. 3 shows an example of oil price forecasting figures. As shown in the top part of Fig. 3, the red curve represents the actual oil prices in the testing part. The black curve represents the predicted oil prices that are calculated by supervised GANs without AS-CWT features, and the blue one represents the predicted price calculated by supervised GANs with AS-CWT features. At the bottom of Fig. 3 shows the predicted error of the two training methods is shown. We can intuitively see that using AS-CWT features can achieve a lower predicted error, which means a better forecasting performance.

## V. Conclusion

In this work, we develop a new forecasting methodology based on supervised GANs with AS-CWT features to forecast short-term crude oil prices. We first use the AS-CWT method to systematically capture the oil prices of different temporal scales by adaptive scales, which can then represent different oil price levels ranging from daily to yearly levels. Then, we develop a supervised GANs to further strengthen the dependence and connection between the input recent oil price and future price. We first compared the nonlinear models (NNs, DBNs) with the traditional economic NF model. The results show that the nonlinear models can outperform the benchmark Random Walk model, and, a comparison between the supervised GANs and the conventional nonlinear models methods (NNs, DBNs) shows that our proposed model can better forecast the values of oil prices (RMSE). Adding the AS-CWT features can make up for the shortage of GANs in forecasting the changing direction of oil price.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K18564 and (A) 17H00983.

## References

[1] J. Alvarez-Ramirez, E. Rodriguez, E. Martina, and C. Ibarra-Valdez, "Cyclical behavior of crude oil markets and economic recessions in the period 1986–2010," *Technological Forecasting and Social Change*, vol. 79, no. 1, pp. 47–58, 2012.

[2] A. Refenes, *Neural networks in the capital markets*. John Wiley & Sons, Inc., New York, USA, 1994.

[3] W. Ongkrutaraksa, "Fractal theory and neural networks in capital markets," 1995, working Paper at Kent State University.

[4] S. Moshiri and F. Foroutan, "Forecasting nonlinear crude oil futures prices," *The Energy Journal*, pp. 81–95, 2006.

[5] S. Wang, L. Yu, and K. K. Lai, "Crude oil price forecasting with tei@i methodology," *Journal of Systems Science and Complexity*, vol. 18, no. 2, pp. 145–166, 2005.

[6] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[8] L. Yu, Y. Zhao, and L. Tang, "A compressed sensing based AI learning paradigm for crude oil price forecasting," *Energy Economics*, vol. 46, pp. 236–245, 2014.

[9] R. Jammazi and C. Aloui, "Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling," *Energy Economics*, vol. 34, no. 3, pp. 828–841, 2012.

[10] M. Tang and J. Zhang, "A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices," *Journal of Economics and Business*, vol. 64, no. 4, pp. 275–286, 2012.

[11] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[12] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[13] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," *arXiv preprint arXiv:1609.03126*, 2016.

[14] E. L. Denton, S. Chintala, R. Fergus *et al.*, "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in neural information processing systems*, 2015, pp. 1486–1494.

[15] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[16] S. Abosedra and H. Baghestani, "On the predictive accuracy of crude oil futures prices," *Energy policy*, vol. 32, no. 12, pp. 1389–1393, 2004.

[17] Z. Luo, J. Chen, T. Takiguchi, and Y. Ariki, "Emotional voice conversion with adaptive scales F0 based on wavelet transform using limited amount of emotional data," *Proc. Interspeech 2017*, pp. 3399–3403, 2017.

[18] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[19] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[21] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, no. 1, 2013.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] C. Li and M. Wand, "Precomputed real-time texture synthesis with Markovian generative adversarial networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 702–716.

[24] D. Kinga and J. B. Adam, "A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[25] T. Nakashika, R. Takashima, T. Takiguchi, and Y. Ariki, "Voice conversion in high-order eigen space using deep belief nets." *in INTERSPEECH*, pp. 369–372, 2013.

[26] Y. Zhao, J. Li, and L. Yu, "A deep learning ensemble approach for crude oil price forecasting," *Energy Economics*, vol. 66, pp. 9–16, 2017.