

해상풍력단지 유지보수 최적화 활용을 위한 풍황 및 해황 장기예측 딥러닝 생성모델 개발

이상훈* · 김대호** · 최혁진*** · 오영진** · 문성빈**

Development of a Deep Learning-based Long-term Prediction Generative Model of Wind and Sea Conditions for Offshore Wind Farm Maintenance Optimization

Sang-Hoon Lee*, Dae-Ho Kim**, Hyuk-Jin Choi***,
Young-Jin Oh** and Seong-Bin Mun**

Key Words : Time-series (시계열), GAN (생성적 적대 신경망), Deep learning (딥러닝), Wind (바람), Wave (파랑),
Offshore wind farm (해상풍력발전단지), Operation & Maintenance (운영 및 유지정비)

ABSTRACT

In this paper, we propose a time-series generation methodology using a generative adversarial network (GAN) for long-term prediction of wind and sea conditions, which are information necessary for operations and maintenance (O&M) planning and optimal plans for offshore wind farms. It is a “Conditional TimeGAN” that is able to control time-series data with monthly conditions while maintaining a time dependency between time-series. For the generated time-series data, the similarity of the statistical distribution by direction was confirmed through wave and wind rose diagram visualization. It was also found that the statistical distribution and feature correlation between the real data and the generated time-series data was similar through PCA, t-SNE, and heat map visualization algorithms. The proposed time-series generation methodology can be applied to monthly or annual marine weather prediction including probabilistic correlations between various features (wind speed, wind direction, wave height, wave direction, wave period and their time-series characteristics). It is expected that it will be able to provide an optimal plan for the maintenance and optimization of offshore wind farms based on more accurate long-term predictions of sea and wind conditions by using the proposed model.

1. 서론

1.1 해상풍력단지 풍황 및 해황 예측의 필요성

원거리 해상에 설치되는 해상풍력단지는 장애물 등의 영향이 적어 육상풍력 대비 풍황 자원적인 측면에서 장점이 뚜렷하며, 입지적인 부분에서도 상대적으로 제약이 적어 대규모 풍력단지 조성에 유리하다. 하지

* 한국전력기술(주) 전력기술원 스마트융합실 (교신저자)

** 한국전력기술(주) 전력기술원 스마트융합실

*** (주)해안해양기술

E-mail : sanghoon@kepco-enc.com

DOI : <https://www.doi.org/10.33519/kwea.2022.13.2.005>

Received : March 31, 2022, Revised : April 29, 2022

Accepted : April 30, 2022

만 이로 인해 해상풍력의 유지보수는 육상풍력에 비해 열악한 해상환경을 추가로 고려하여야 한다.

해상풍력의 유지보수 작업은 보통 선박을 이용하므로 파고, 풍속 등 기상 상태에 따른 선박 운항 및 해상작업 가능 여부에 대한 예측과 정비접근 제한으로 인한 풍력발전기 비가동률의 최소화 등을 고려하여 최적화 및 최적계획을 수립하여야 한다. 유지보수 정비 계획에 고려되는 주요 기상 정보는 유의파고, 풍향 그리고 풍속이다. 유지보수를 위한 선박은 일반적으로 유의파고 1.5 m 및 풍속 10 m/s 이상이면 풍력단지 접근이 제한되는 경우가 많으며, 풍속과 유의파고 데이터를 통해 유지보수를 위한 풍력터빈 접근이 가능한 일정을 예측할 수 있다 [1]. 대규모 해상풍력단지의 경우 기상 정보를 바탕으로 이루어지는 인력, 자원, 작업 등의 최적 배치를 통한 정비 활동으로 비용을 절감할 수 있고, 동시에 터빈 비가동 시 풍황 정보기반 발전 손실량 예측을 통해 유지보수 작업계획을 수립할 수 있다. 따라서 풍향 및 해황 기상 상태의 장기적인 예측은 해상풍력단지의 유지보수 최적화 및 최적계획 수립을 위해서 반드시 필요하다.

1.2 해상 작업가능일수 평가 연구 동향

해상작업 가능 기간 산정에 대한 별도의 지침은 없으며 선박의 운항 가능 여부는 해상교통안전법에 따라 폭풍, 파랑, 해일주의보 및 경보와 시계 제한과 같은 제약이 있고 우리나라에서는 해상작업 가능 기간 산정이 주관적 판단에 따르는 경향이 있다.

해상 건설공사에서 적용되고 있는 해상작업 작업일수 산정 또한 최근까지도 약 50년 전에 제안 및 사용된 기상현상에 따른 작업한계 기반 산정 방법을 이용해 왔으며, 이는 해상 기상 조건 중 중요한 파랑과 같은 요소가 고려되지 않은 방법이다 [2]. 육상에 비하여 열악한 해상이라는 특성을 감안하고 해상작업 시 안전사고와 작업중단에 큰 영향을 끼치는 파랑을 해상작업일수 산정에 고려하려는 연구 및 필요성이 제기되었다 [3, 4]. 해양수산부에서는 해상공사의 작업환경 특성을 감안하여 파랑 등의 해상 기상 조건을 고려한 비작업일수 산정방식을 Fig. 1과 같이 제시하였다 [5]. 파랑을 고려한 해상 작업일수 산정을 위해서는 파업대상지에서의 10년 이상 장기 파랑 자료를 통계분석을 통하여 주기 및 파고의 세기에 따른 계급별 출현빈도를 분석한 후 주공정의 작업한계파고를 적용하여 파랑에 의

한 비작업일수를 산정하도록 하고 있다.

해상작업 가능 기간 산정을 위하여 확률 모형을 적용하는 시도들도 있었으며 이는 장기 파랑자료의 수치모의를 통하여 역추산된 유의파고, 침두주기 자료로부터 장기 유의파고 확률 분포를 도출하는 방법을 사용하였다. 이와 같은 확률모형을 이용한 방법들은 파랑자료 자체에 집중하여 알려진 단기 및 장기 파랑의 파고 분포함수의 모수를 추정하는 방식을 사용하였다. 확률모형을 이용한 최근 연구에서는 단기 확률모형을 유의파고 시계열 자료에 대해 선회 적분하는 과정을 통해 개별 파랑이 고려된 장기 파고 분포함수를 도출하여 해상작업 가능 기간을 산정하였다 [6].

기존 방법들은 10년 이상의 기상자료 관측치 또는 수치모델에 의해 추산된 과거 데이터를 활용하여 통계처리 또는 특정 데이터의 알려진 확률모형을 적용한 것이다. 풍력단지 설계단계에서 필요한 발전량 및 파고의 최댓값 등의 예측은 풍속이나 파고의 단변량 변수를 활용한 단순 통계치 분석 또는 기존 방법들의 활용으로 가능하며 이를 적용하는 것은 큰 문제가 없다. 하지만 해상작업 가능 기간의 경우는 풍속과 파고 등의 복잡한 관계를 동시에 고려하여야 하므로 기존 방법들의 적용에는 어려움이 있다. 특히 해상풍력단지의 유지보수 최적화 및 최적계획 수립을 위해 과거 데이터를 활용한 기존 방법들을 적용하는 것은 월간 작업가능일, 월간 연속 시간 동안 작업의 중요성, 풍력단지의 지역 특성 등을 복합적으로 고려하지 못하며, 통계적인 유의성을 보장하기에는 가용 데이터의 수가 많이 부족하다.

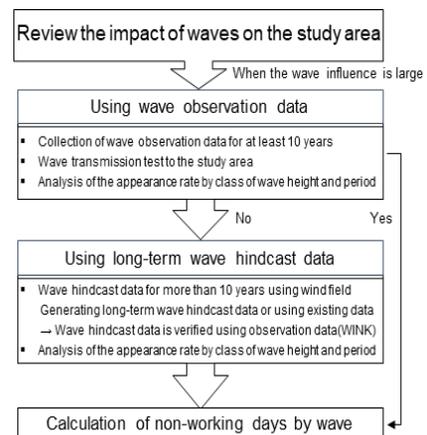


Fig. 1 Procedure diagram for calculating the number of non-working days by wave [5]

또한 풍황 또는 해황 각각의 독립적인 단변량 분포 추정 및 통계처리로는 해상환경의 서로 상관관계가 복잡하게 얽혀있는 풍속, 풍향, 유의파고, 파향, 유의파주기 와 같은 특성들을 동시에 고려할 수 없다는 한계점이 있다.

1.3 연구의 목적

풍황 및 해황 장기예측을 위하여 시계열 간의 시간의존성, 다변수 간의 상관성, 해상풍력단지의 공간적 특성, 예측하고자 하는 대상년도의 특성 등을 동시에 고려할 수 있는 새로운 접근법이 필요하다.

이를 위해 다변수간의 복잡한 상관관계를 데이터에 기반하여 상관관계를 가진 데이터들의 분포를 동시에 모사할 수 있는 딥러닝 생성모델(Generative Model) 중 하나인 GAN 모델을 이용한 시계열 데이터 생성 방법론을 검토해보고 생성된 확장 시계열 데이터를 바탕으로 풍황 및 해황의 조건을 동시에 고려한 작업가능일 산출의 가능성을 검토해보고자 한다.

본 논문의 구성은 다음과 같다. 우선 2장에서 시계열 데이터 생성모델로서의 GAN 아키텍처가 적용된 선행연구를 살펴보고 월별 시계열 데이터 생성이 가능한 GAN 기반의 수정된 모델을 제안한다. 3장에서는 제안된 GAN 기반의 시계열 생성모델의 적용성을 검토하고 결과를 분석한다. 4장에서는 본연구의 결론과 추후 연구 방향에 관한 내용을 제시한다.

2. 풍황 및 해황 장기예측에 대한 검토

2.1 GAN 선행연구 검토

생성모델은 주어진 데이터를 바탕으로 데이터 자체가 가지는 패턴과 구조를 비지도학습(Unsupervised Learning)을 통하여 데이터의 분포를 따르는 유사한 데이터를 생성하는 모델이다. 이러한 생성모델에는 여러 가지 방식이 있는데, 크게 데이터의 확률분포를 명시적으로 추정하여 적용하는 명시적 방법(Explicit Density)과 학습 데이터의 분포를 모르는 상태에서 적용하는 암묵적 방법(Implicit Density)으로 나누어 볼 수 있다. 명시적 방법에는 학습 데이터의 분포를 직접적으로 구하는 방법(Tractable Density)이 있고, 이러한 분포를 단순히 추정하는 방법(Approximate Density)이 있다. 본 연구에서는 암묵적 방법의 생성모

델 중에서도 최근 가장 많이 사용되고 있는 딥러닝 기반의 생성모델인 GAN 모델의 적용을 검토하고자 한다 [7].

GAN을 이용한 연구는 이미지 생성 및 변환 분야에서 활발하게 진행되고 있다. GAN을 이용한 이미지 생성 기술은 컨볼루션 구조를 이용한 DCGAN(Deep Convolutional Generative Adversarial Network), 이미지 데이터 분포가 아닌 손실함수 자체의 분포에 초점을 맞추는 BEGAN(Boundary Equilibrium Generative Adversarial Network), 에너지 기반의 생성모델인 EBGAN (Energy-Based Generative Adversarial Network), Wasserstein Distance를 도입하여 두 확률분포 간의 사이를 측정하는 WGAN(Wasserstein Generative Adversarial Network)등 다양하게 연구 및 발전되고 있다 [8-11].

생성모델을 이용하여 시계열 간의 시간의존성(Temporal Dynamics)을 유지하는 시계열 데이터를 생성하는 것은 상당히 어려운 문제이다. 시간의존성은 단위 시간에 따라 바뀌는 변수 간의 복잡한 변화를 의미한다.

특히 이미지 생성 및 변환 분야에 특화되어있는 GAN의 경우 시계열 생성 분야에 적용한 사례는 상대적으로 많지 않다. 몇몇 연구들에서 시계열 데이터 생성을 위하여 순서가 의미가 있는 시퀀스 데이터에 적합한 순환신경망(Recurrent Neural Networks, RNN)을 이용하여 GAN을 구성하는 시도가 있었다. 이는 주로 생성자(Generator)와 판별자(Discriminator)의 구성요소에 순환신경망 네트워크를 사용하고 기본적인 GAN 모델의 프레임워크를 유지하는 데 중점을 두었다.

C-RNN-GAN(Continuous Recurrent Neural Networks with Adversarial Training) 또한 생성자 및 판별자에 순환신경망의 일종인 LSTM(Long Short-Term Memory) 네트워크를 사용하고 기본적인 GAN 모델의 프레임워크를 시계열 데이터 생성에 직접 적용하여 음악 데이터를 합성하였다 [12]. 데이터는 노이즈 벡터와 이전 시간 단계에서 생성된 출력 데이터를 입력으로 사용하여 반복적으로 생성된다. 연속 시퀀스 데이터로 훈련하며 음악 영역에서는 높은 유연성과 표현력 유지에 사용될 수 있다는 것을 보여주었고, 이는 길고 연속적인 시계열 생성이 가능하다는 장점이 있다.

RCGAN(Recurrent Conditional GAN)도 유사한 방식으로 기본적인 GAN 모델의 프레임워크에 순환신경

망을 적용하고 이전 시간 단계에서 생성된 출력 데이터에 대한 컨디셔닝 대신 입력에 대해 추가적인 컨디셔닝을 하는 것과 같이 아키텍처의 일부 변경을 시도하여 단순한 사인파나 의료데이터를 생성하였다 [13].

C-RNN-GAN 및 RCGAN에서 살펴본 바와 같이 GAN의 생성자와 판별자에 단순히 순환신경망을 적용하는 것으로는 시계열 간의 시간의존성을 유지할 수 없으며 이들 구조는 특정 도메인에서만 작동한다는 단점이 있다.

TimeGAN(Time-series GAN)은 이와 같은 문제점을 해결하기 위해 데이터의 분포를 모사하는 비지도학습 기반의 이진(Binary) 적대적 피드백에 시간의존성을 학습할 수 있는 지도학습을 추가로 도입하였다 [14]. 이를 효과적으로 수행하기 위해 임베딩 네트워크를 활용하여 잠재 공간(Latent Space)을 생성하고 잠재 공간 내에서 시계열 간의 시간의존성을 지도학습을 통해 배울 수 있도록 하였다.

Fig. 2는 TimeGAN의 기본적인 구조를 도식화한 것이다. TimeGAN은 특정 도메인뿐만 아니라 다양한 도메인의 데이터 생성으로 일반화할 수 있고, 생성된 시계열 데이터는 시계열 간의 시간의존성을 유지할 수 있다. 하지만 TimeGAN을 활용하여 생성한 시계열 데이터는 그 시계열 데이터의 레이블, 즉 데이터의 카테고리 정보를 알 수 없는 랜덤 샘플링 데이터라는 한계가 있다.

풍황 및 해황의 예보를 월 단위로 제어하며 생성하기 위해서는 TimeGAN에 조건설정이 필요하다는 것을 알 수 있다. 이를 위해 본 연구에서는 TimeGAN 프레임워크의 생성자와 판별자에 추가 정보를 제공하는 Conditional GAN의 아이디어를 접목하여 시계열 간의 시간의존성을 유지하며 월 단위로 제어가 가능한 풍황 및 해황 시계열 데이터 생성방안인 Conditional TimeGAN을 제안하고자 한다 [15].

2.2 Conditional TimeGAN 개념 제안

Conditional TimeGAN은 기본적으로 TimeGAN의 구성요소를 그대로 사용하며, 총 4가지 신경망 네트워크 구성 요소인 임베딩 층(Embedding Layer), 복구 층(Recovery Layer), 생성자(Generator) 그리고 판별자(Discriminator)로 구성된다 (Fig. 3 참조).

임베딩 층(e_x)은 입력 데이터를 저차원 공간으로 매

핑하는 기능을 수행하는 신경망 층으로 LSTM이나 GRU(Gated Recurrent Unit)와 같은 순환신경망으로 구성되며, x_t 는 t 시점의 시계열 입력 데이터, h_t 는 t 시점의 저차원 공간에서의 특성(Feature)이다.

$$h_t = e_x(h_{t-1}, x_t) \quad (1)$$

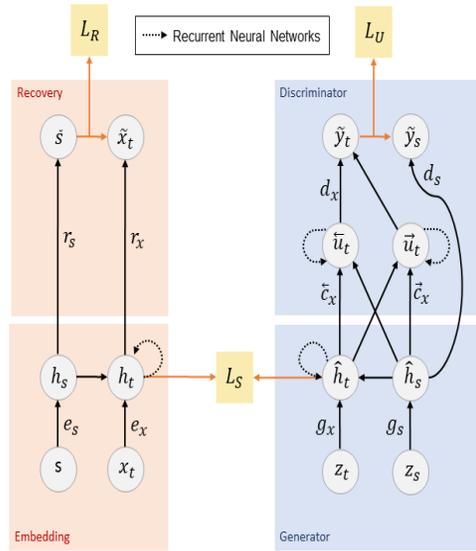


Fig. 2 TimeGAN architecture [14]

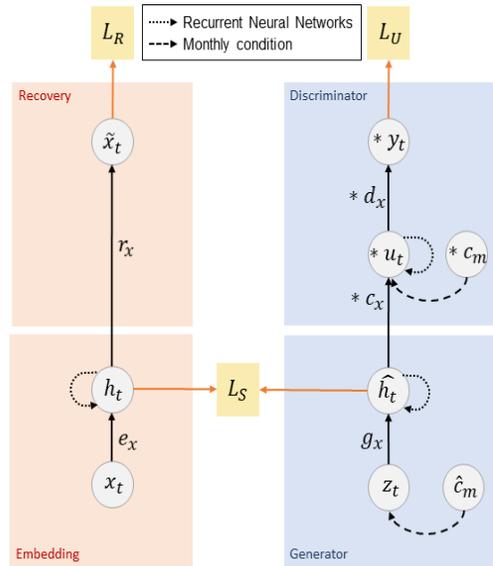


Fig. 3 Conditional TimeGAN architecture (proposed)

복구 층(r_x)은 임베딩 층을 통해 매핑된 저차원 공간 특성을 원본 데이터로 복원하는 기능을 수행하는 신경망 층으로 LSTM이나 GRU와 같은 순환신경망으로 구성되며, \tilde{x}_t 는 t 시점의 시계열 입력 데이터의 복원된 시계열 데이터이다.

$$\tilde{x}_t = r_x(h_t) \quad (2)$$

생성자(g_x)는 랜덤 노이즈 벡터로부터 저차원 공간에서의 가짜(Fake) 데이터를 생성하는 기능을 수행하는 신경망 층으로 LSTM이나 GRU와 같은 순환신경망으로 구성되며, \hat{h}_t 는 t 시점의 저차원 공간에서의 가짜 생성 데이터의 특성, z_t 는 랜덤 노이즈 벡터, \hat{c}_m 은 임의의 월 정보를 인코딩한 벡터이다.

$$\hat{h}_t = g_x(\hat{h}_{t-1}, z_t, \hat{c}_m) \quad (3)$$

판별자(d_x)는 저차원 공간에서 임베딩 층을 통해 출력된 실제 특성과 생성자에 의해 생성된 특성 간 확률분포의 적대적 학습 기능을 수행하는 신경망 층으로, c_x 는 순환신경망으로 구성되어 있으며, u_t 는 t 시점의 저차원 공간에서의 적대적 학습의 판단을 위한 특성, c_m 은 월 정보를 인코딩한 벡터, y_t 는 0과 1 사이의 실제 또는 가짜 생성 데이터일 확률값이다. $\hat{\cdot}$ 기호는 가짜 생성 데이터 관련임을 나타내는 기호이며 Fig. 3상의 * 기호는 실제 데이터와 가짜 생성 데이터 2가지 경우를 모두 표현한 것이다.

$$u_t = c_x(h_t, u_{t-1}, c_m), \hat{u}_t = \hat{c}_x(\hat{h}_t, \hat{u}_{t-1}, \hat{c}_m) \quad (4)$$

$$y_t = d_x(u_t), \hat{y}_t = \hat{d}_x(\hat{u}_t) \quad (5)$$

Conditional TimeGAN의 손실함수(Loss Function)는 크게 3가지로 구성되며, 첫째는 저차원의 임베딩 공간에서 이루어지는 학습을 위하여 입력된 데이터를 다시 복원하였을 때 가역적으로 최대한 손실 없이 구현해 낼 수 있는 함축의 기능을 도와주는 복원손실(Reconstruction Loss) L_R ,

$$L_R = E_{X_{1:T} \sim p} [\sum_t \|x_t - \tilde{x}_t\|_2] \quad (6)$$

둘째는 학습된 임베딩 공간 안에서 컨디셔닝된 적대적

피드백을 통한 실제 데이터의 분포를 모사하는 기능을 도와주는 비지도 학습손실(Unsupervised Loss) L_U ,

$$L_U = E_{X_{1:T} \sim p} [\sum_t \log y_t] + E_{X_{1:T} \sim \hat{p}} [\sum_t \log(1 - \hat{y}_t)] \quad (7)$$

셋째는 실제 데이터를 감독으로 사용하여 단계적 조건부 확률 분포를 추가로 학습하여 $p(H_t | C_m, H_{1:t-1})$ 과 $\hat{p}(H_t | \hat{C}_m, H_{1:t-1})$ 의 차이를 줄이는 데 도움을 주는 지도 학습손실(Supervised Loss) L_S ,

$$L_S = E_{X_{1:T} \sim p} [\sum_t \|h_t - g_x(h_{t-1}, z_t, \hat{c}_m)\|_2] \quad (8)$$

상기의 세 가지 손실함수는 유기적으로 연결되어 공동으로 학습됨으로써 잠재공간의 효율성을 극대화하고 시계열 간의 컨디셔닝된 단계적 조건부 분포를 용이하게 캡처하여 시계열 간의 시간의존성을 유지하는 실제 시계열 데이터와 유사한 데이터를 월 단위 조건 설정으로 제어하며 생성할 수 있게 한다.

3. Conditional TimeGAN 적용성 검토

3.1 입력 데이터 준비

파랑은 풍속, 풍향과 밀접한 관련이 있고 파랑에 내재한 불규칙성을 고려할 때 풍속, 풍향, 유의파고, 파향, 유의파 주기는 서로 영향을 미치며 복잡한 상관관계를 가진다. 딥러닝 기반의 GAN 알고리즘을 적용하기 위하여 풍속, 풍향, 유의파고, 파향, 유의파 주기를 입력으로 사용한다. 우리나라의 경우 2020년을 기준으로 기상청 20개소, 국립해양조사원 6개소, 해양수산부 14개소에서 지속적으로 실시간 파랑관측을 수행하고 있으나, 동해가스전 위치에서의 파랑관측자료는 부재하여 해상풍 자료와 수치모델에 기초한 재현자료를 생산하여 사용한다.

동해가스전 위치에서의 해황 과거 재현자료(1979년에서 2020년까지)를 생성하기 위해서 바람장을 이용한 파랑 추산 수치 실험을 수행하였으며 수치모형실험의 입력으로 사용한 바람장은 공간적인 해상도가 뛰어난 일본기상청(Japan Meteorological Agency, JMA)의 예보 바람장을 기본으로 사용되되, 일본기상청의 자료가 없는 2002년 7월 이전은 해양수산부에서 운영하고 있는 전국파랑관측자료 제공시스템(Wave Information



Fig. 4 Train on generated, test on original (TGTO)

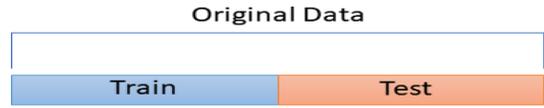
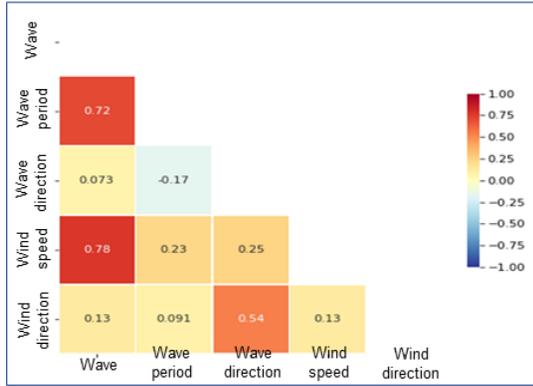
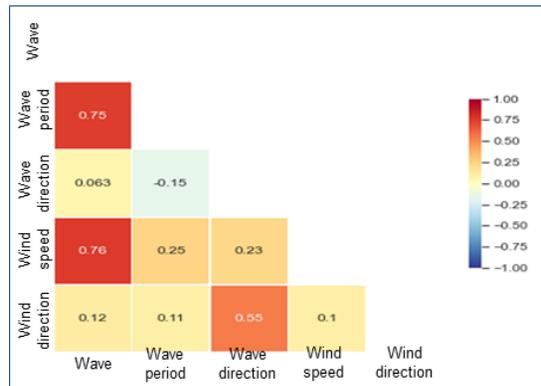


Fig. 5 Train on original, test on original (TOTO)



(a) Original



(b) Generation

Fig. 6 Heat map

Network of Korea, WINK)을 통하여 제공되고 있는 해상풍 재분석 자료(통칭 KIOST 바람장)를 사용하였다.

수치모델은 SWAN(Simulation WAVes Nearshore)모델을 이용하였고 유의파고, 유의파 주기, 파향 등을 한 시간 간격으로 추산하였으며 추산된 자료의 검증은 기상청(Korea Meteorological Agency, KMA) 파랑 관측자료를 활용하여 수치 모형 실험의 재현성을 확인하였다.

3.2 결과분석

본 연구에서는 동해가스전 위치에서의 42년(1979년에서 2020년까지) 동안의 풍향 및 해황의 과거 재분석 자료를 전처리하여 GAN기반의 데이터 생성모델을 훈련 및 구축하였다. 풍속, 풍향, 유의파고, 파향, 유의파 주기를 입력 데이터로 사용하여 월 레이블별 풍속, 풍향, 유의파고, 파향, 유의파 주기의 1일 단위, 1시간 간격의 연속된 시계열 약 1000년치를 생성하였다.

시계열 예측은 생성된 시계열 데이터가 실제 시계열 데이터를 잘 나타내는지 여부를 정량적으로 평가하는데 사용될 수 있다. 이는 “Train on Generated, Test on Original (TGTO)”과 “Train on Original, Test on Original (TOTO)” 두 가지 방법을 통해 산출된 평가 지표를 비교함으로써 가능하다. TGTO는 생성된 시계열

데이터 세트를 사용하여 LSTM 네트워크 기반 시계열 예측 모델을 학습하고 각 입력 시퀀스에 대한 다음 단계의 시퀀스를 예측하도록 한다. 학습이 완료된 모델에 실제 시계열 데이터 세트를 입력으로 하여 학습된 모델을 평가한다 (Fig. 4 참조).

TOTO는 실제 시계열 데이터 세트를 학습용과 테스트용으로 나누고, 학습용 데이터를 사용하여 LSTM 네트워크 기반 시계열 예측 모델을 학습 및 다음 단계의 시퀀스를 예측하도록 하고, 학습이 완료된 모델에 테스트용 데이터 세트를 입력으로 하여 학습된 모델을 평가한다 (Fig. 5 참조). 평가 지표로 MAE(Mean Absolute Error)를 사용, 점수가 낮을수록 성능이 더 좋음을 의미하며 TGTO 점수가 TOTO 점수에 최대한 가깝게 낮은 점수를 얻는 것이 목표이다. 생성된 시계열 데이터를 사용하여 산출한 TGTO 점수는 0.031, TOTO 점수는 0.020으로 충분히 낮은 점수를 보인다고 할 수 있다.

Fig. 6은 실제 데이터와 생성된 데이터의 특성 상관관계를 나타낸 히트맵(Heat Map)으로 상관관계가 높을수록 진한 색으로 표현되며 각각의 박스 안에는 상관계수가 표현된다. 상관계수는 두 특성 사이의 통계적 관계를 표현하기 위해 특정한 상관관계의 정도를 수치적으로 나타낸 계수로서 값의 범위는 -1과 +1 사이이다. 양의 상관관계일 경우 (+), 음의 상관관계일

경우 (-)값이 나타난다. 실제 데이터의 경우 유의파고와 풍속, 유의파주기는 상관계수 0.7 이상의 강한 상관관계를 보이며 유의파주기와 파향은 -0.17로 약하지만 음의 상관관계 경향을 보인다. 생성된 데이터의 경우 유의파고와 풍속, 유의파주기는 상관계수 0.7 이상의 강한 상관관계, 유의파주기와 파향은 -0.15로 약하지만 음의 상관관계 경향을 보이며 이는 실제 데이터의 상관관계와 유사하다. 나머지 특성들도 생성된 데이터와 실제 데이터의 상관계수를 비교해 보면 그 값이 큰 차이가 없으며 이는 특성 간 상관관계를 상당히 유사하게 표현하는 것으로 볼 수 있다.

Fig. 7은 방위별 파향 출현빈도와 파향별 파고계급 빈도, Fig. 8은 방위별 풍향 출현빈도와 풍향별 풍속계급 빈도를 장미도로 도시하여 월별로 비교한 것이다. 이러한 장미도는 특정 지점에서 어느 기간 동안 각 방위별, 방향성분을 가진 기상 데이터의 빈도를 방사 모양의 그래프로 나타낸 것으로, 출현 빈도의 백분율(%)을 각각의 방향성분에 대응하는 방위판 위에 방위선 길이로 표현한 것이다. 파랑 및 바람 장미도는 방향성분과 크기로 이루어진 벡터 형태의 기상데이터인 파랑과 바람을 그림의 형태로 시각화하여 파향 및 풍향에 따른 파고와 풍속의 빈도와 분포를 한눈에 알아보기 쉽게 한다. 1월에서 12월의 파랑 및 바람 장미도를 살펴보면 일부 빈도의 차이를 보이는 월도 있지만 대체로 파랑과 바람이 주로 분포되어 있는 방향과 그 세기별 출현빈도가 대부분 일치하며 이는 실제 데이터와 생성된 데이터의 유의파고와 파향, 풍속과 풍향의 결합데이터의 분포 특성이 상당히 유사한 경향을 보이는 것으로 볼 수 있다.

Fig. 9는 1일(1시간 간격) 기준으로 생성된 파고, 풍속 시계열 데이터를 이용하여 작업 가능 시간대(ex. 오전 7시와 오후 7시 사이)의 연속되는 일정시간(ex. 8시간연속) 동안 파고 1 m 이하이고 풍속 14 m/s 이하인 작업가능일수를 통계분석 하여 월별 작업가능일수를 박스플롯으로 비교한 것이다. Fig. 7과 Fig. 8에서 살펴본 바와 같이 12월에서 2월까지의 겨울철로 파고와 풍속이 높은 경향이 있어 작업가능일이 적으며, 6월에서 8월까지는 여름철로 파고와 풍속이 낮은 경향이 있어 작업가능일이 많게 예측된다. 전체적인 월별 작업가능일수는 실제 데이터에 비하여 다소 보수적인 예측치를 보이는 월도 있지만 대체로 경향은 유사하게 추정하는 것으로 나타나며, 생성된 데이터는 실제 데이터보다 표본의 수가 많아 사분위수 범위가 실제 데이터보다 좁게 나타난다.

고차원 데이터의 경우 특성 간의 분포 및 관계를 수리적으로 분석하기 어려우며, 이를 효과적으로 분석하고 핵심 정보를 선정할 수 있는 시각화 방법으로 PCA(Principal Component Analysis)와 t-SNE(Stochastic Neighbor Embedding) 알고리즘이 있다. Fig. 10은 PCA 및 t-SNE 알고리즘을 사용하여 생성된 데이터와 실제 데이터 간의 통계적 분포를 시각화하여 보여주고 있다. 이는 높은 차원의 복잡한 데이터를 2차원에 차원 축소하는 방법으로 낮은 차원 공간으로의 시각화에 주로 사용하며 높은 차원을 낮은 차원으로 축소하는 과정에서 비슷한 구조끼리 데이터를 정리한 상태로 보여줌으로 데이터 간의 유사도를 쉽게 파악할 수 있다. 실제 데이터와 생성된 데이터의 타점 분포를 통해 해당 데이터의 매니폴드를 시각적으로 확인할 수 있고 이것이 높은 비율로 일치할수록 실제 데이터를 잘 표현하고 있다고 볼 수 있다. Fig. 10의 타점 분포를 확인해 보면 실제 데이터와 Conditional TimeGAN으로 생성한 데이터 간의 통계적 분포가 거의 일치한다는 것을 확인할 수 있다.

4. 결론

본 연구에서는 해상풍력단지 유지보수 최적화 및 최적계획을 위하여 필요한 풍향 및 해황의 장기 예측 추론치의 산출 방법을 제안하였다. 기존에 사용되던 방법은 10년 이상 장기파랑 과거 자료를 토대로 통계 분석 또는 확률모형을 적용하는 방법으로, 풍속 또는 유의파고와 같이 단변량 데이터만을 고려하여 예측한다. 하지만 이는 해상풍력단지의 유지보수를 위한 복잡한 해상 상황을 복합적으로 고려할 수 없다. 이에 부족한 데이터의 확장 및 풍속, 풍향, 유의파고, 파향, 유의파 주기의 서로 복잡한 비선형적 상관관계를 동시에 고려할 수 있는 딥러닝 기반 시계열 데이터의 생성 모델을 통한 장기예측치를 추론하였다. 시계열 간의 시간의존성을 보장하고 원하는 월 레이블의 데이터 생성이 가능한 GAN 기반의 Conditional TimeGAN 모델을 제시하였고 이를 이용하여 풍속, 풍향, 유의파고, 파향, 유의파 주기 5개의 특성을 출력으로 하는 월 단위의 약 1000년 치 시계열 데이터를 생성하였다.

실제 데이터와 생성 데이터의 방향별 유사성을 확인하기 위해 파랑, 바람장미도를 사용하여 방위별 및 계급별 출현빈도의 유사성을 확인하였고, 고차원 데이터의 저차원에서의 시각화 알고리즘인 PCA와 t-SNE를 사용

하여 통계적 유의성을 확인하였다. 생성된 데이터를 이용하여 풍속과 유의파고를 특정 시간대별 기준값을 한 계로 연속 작업 가능 시간을 고려한 작업가능일수를 산출하여 실제 데이터와 비교하여 유사성을 확인하였다.

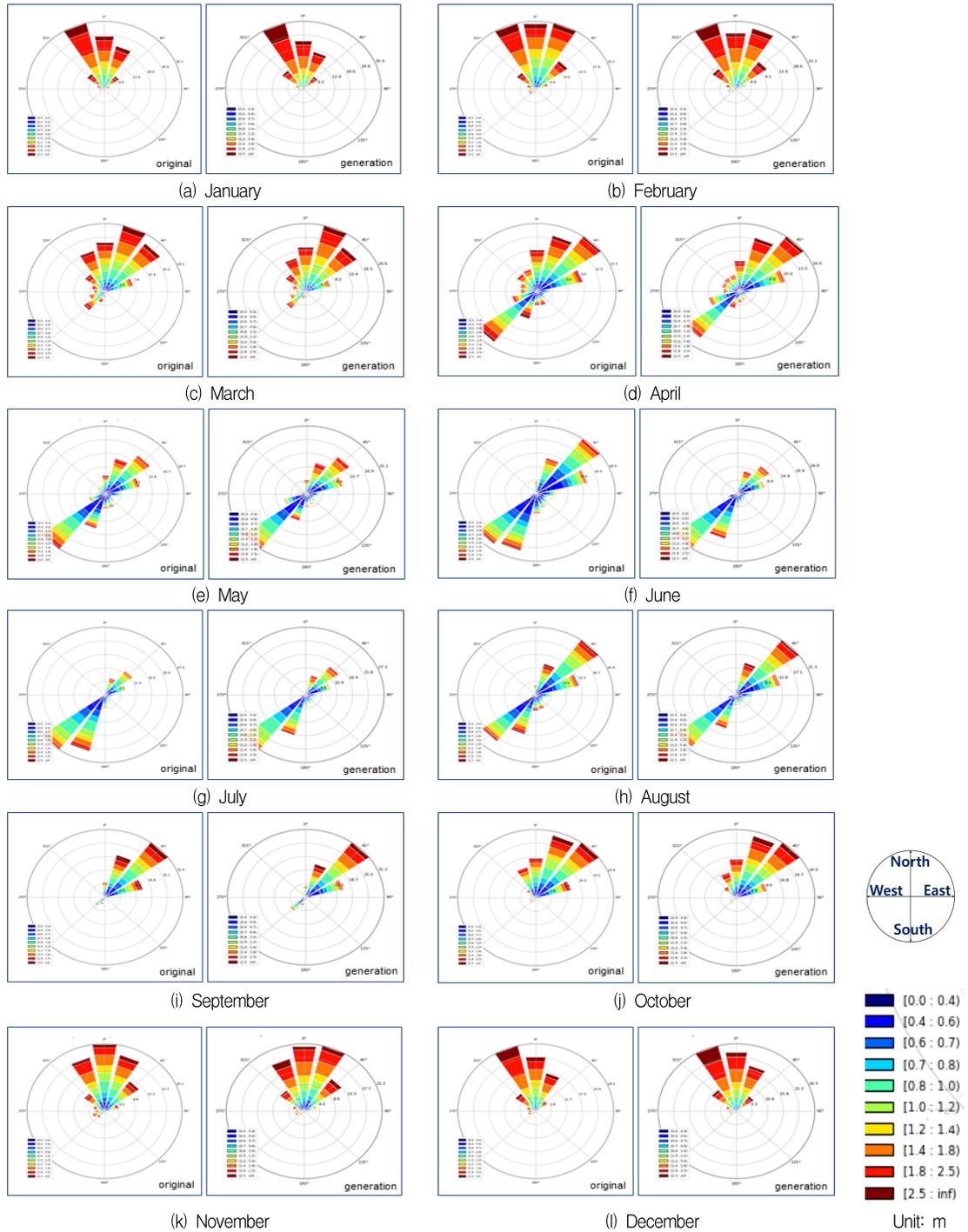
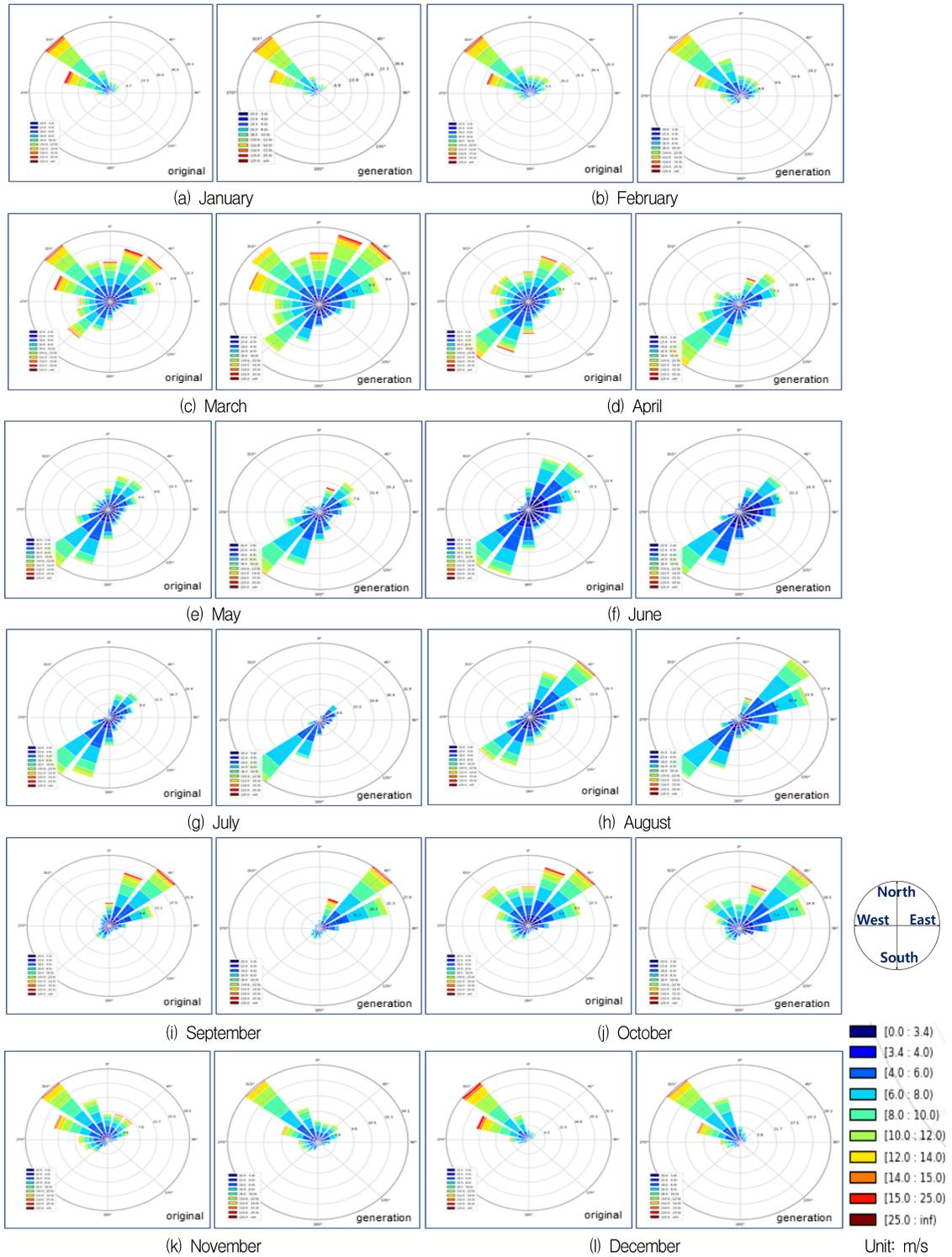


Fig. 7 Wave rose (left: original, right: generation)



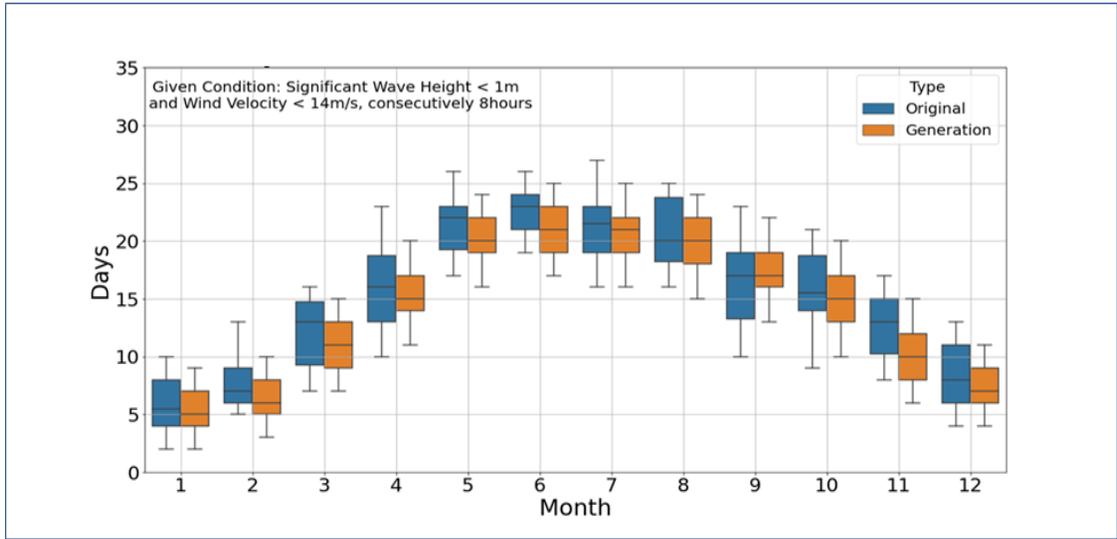


Fig. 9 Monthly workable days

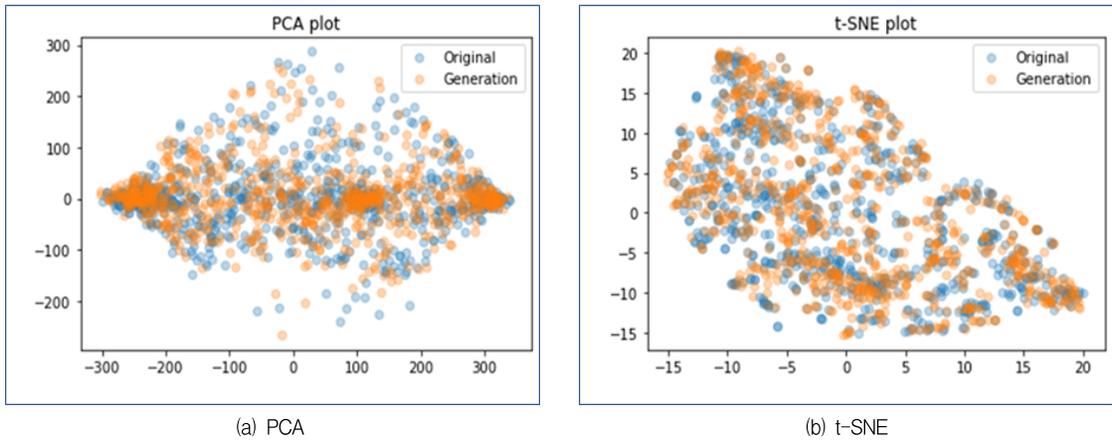


Fig. 10 PCA and t-SNE visualization

본 논문은 시간의존성을 보존하는 시계열 생성모델에 컨디션링을 추가하여 원하는 월 단위별 시계열 데이터 생성과 같은 조건부 확률분포 제어의 가능성에 관하여 연구하였다. 향후 풍력단지의 공간적 특성, 예측하고자 하는 대상년도의 특성 등의 유지보수 관점에서 추가로 필요한 정보들의 조건부 확률분포를 제어하여 다양한 조건을 고려한 시계열 데이터를 생성할 수 있기를 기대한다.

후기

본 연구는 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구입니다(과제번호: 20203010020050).

참고문헌

- [1] Sang, M. G., Lee, N. K., Shin, Y. H., Lee, C., and Oh, Y. J., 2021, "An Optimization Model for O&M Planning of Floating Offshore Wind Farm using Mixed Integer Linear Programming," Journal of the Korea Society of Computer and

- Information, Vol. 26, No. 12, pp. 255-264 (in Korean).
- [2] Ro, H.Y., 1969, "Problems on design and construction used heavy equipments," Journal of Korean Society of Civil Engineers, Vol. 17, No. 1, pp. 54-69 (in Korean).
- [3] Choi, S. H., Back, J. D., Jeong, W. M., and Choi, H. J., 2019, "A study on the estimation of the number of working days for harbor construction project considering wave condition," Journal of Korean Society of Coastal Disaster Prevention, Vol. 6, No. 2, pp. 71-82 (in Korean).
- [4] Back, J. D., Jeong, W. M., Lee, J. I., Choi, H. J., and Kim, M. K., 2021, "A Study on the Method of Calculating the Number of Working Days for Harbor Construction using the Wave of Ulsan New Port," Journal of Korean Society of Coastal and Ocean Engineers, Vol. 33, No. 2, pp. 80-91 (in Korean).
- [5] Ministry of Oceans and Fisheries, 2020, Report on the design practice for harbor construction project (in Korean).
- [6] Choi, S. H. and Cho, Y. J., 2019, "Development of a Probabilistic Model for the Estimation of Yearly Workable Wave Condition Period for Offshore Operations - Centering on the Sea off the Ulsan Harbor," Journal of Korean Society of Coastal and Ocean Engineers, Vol. 31, No. 3, pp. 115~128 (in Korean).
- [7] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., and Bengio, Y., 2014, "Generative adversarial nets," Advances in neural information processing systems 27.
- [8] Radford, A., Metz, L., and Chintala, S., 2015, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434.
- [9] Berthelot, D., Schumm, T., and Metz, L., 2017, "Began: Boundary equilibrium generative adversarial networks," arXiv preprint arXiv:1703.10717.
- [10] Zhao, J., Mathieu, M., and LeCun, Y., 2016, "Energy-based generative adversarial network," arXiv preprint arXiv:1609.03126.
- [11] Arjovsky, M., Chintala, S., and Bottou, L., 2017, "Wasserstein generative adversarial networks," In International conference on machine learning, pp. 214-223, PMLR.
- [12] Mogren, O., 2016, "C-RNN-GAN: Continuous recurrent neural networks with adversarial training," arXiv preprint arXiv:1611.09904.
- [13] Esteban, C., Hyland, S. L., and Rätsch, G., 2017, "Real-valued (medical) time series generation with recurrent conditional gans", arXiv preprint arXiv:1706.02633.
- [14] Yoon, J. S., Jarrett, D., and Van der Schaar, M., 2019, "Time-series Generative Adversarial Networks", Advances in Neural Information Processing Systems, 32.
- [15] Mirza, M. and Osindero, S., 2014, "Conditional Generative Adversarial Nets", arXiv preprint arXiv:1411.1784.