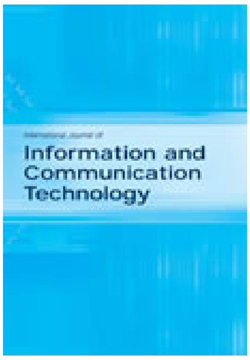


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Regional wind speed prediction model based on graph attention network and meteorological coupled equations

Yuechao Zhang

DOI: 10.1504/IJICT.2025.10074946

Article History:

Received:	07 September 2025
Last revised:	16 October 2025
Accepted:	17 October 2025
Published online:	17 December 2025

Regional wind speed prediction model based on graph attention network and meteorological coupled equations

Yuechao Zhang

Long Yuan (Beijing) New Energy Engineering Technology Co., Ltd.,
Beijing, 100034, China
Email: yuechaoz88@163.com

Abstract: Aiming at the challenges of spatio-temporal nonlinearity and physical consistency in regional wind speed prediction, this paper proposes graph attention network with physical constraints, a new model that fuses the graph attention network and meteorological equations. The model dynamically captures the complex relationship between meteorological stations through the graph attention mechanism and jointly optimises the loss function with the horizontal momentum equation as the physical constraint. Experiments based on high-resolution data in the Beijing-Tianjin-Hebei region from 2018–2021 show that the root mean square error and mean absolute error of graph attention network with physical constraints in 24-hour forecasts are 1.52 m/s and 1.11 m/s, respectively, which are reduced by 11.1% and 11.9% compared to the optimal baseline, and the R^2 is improved to 0.948. Its excellent performance in extreme events provides a new paradigm for high-precision, interpretable weather prediction.

Keywords: wind speed prediction; graph attention networks; physically informed machine learning; coupled meteorological equations; spatio-temporal prediction.

Reference to this paper should be made as follows: Zhang, Y. (2025) 'Regional wind speed prediction model based on graph attention network and meteorological coupled equations', *Int. J. Information and Communication Technology*, Vol. 26, No. 47, pp.71–88.

Biographical notes: Yuechao Zhang received his Master's degree from Beijing University of Technology in 2010. He currently serves as the Director of the Data Information Center at Longyuan (Beijing) New Energy Engineering Technology Co., Ltd. His research focuses on artificial intelligence, big data predictive technologies, computing infrastructure, data resource governance and value mining.

1 Introduction

As a central pillar of the global energy mix transition, its efficient development and utilisation is highly dependent on the accurate prediction of future wind conditions, especially near-surface or hub-height wind speeds (Council, 2021). Highly accurate regional wind speed prediction is not only a key prerequisite for the power prediction of wind farms, power grid scheduling and stable operation, but also an important link in aviation and navigation safety, weather warning and climate research. However, as a highly nonlinear chaotic system, the evolution of the atmospheric system is governed by the complex interactions of multi-scale and multi-physical processes (e.g., thermal, dynamical, radiative, etc.), which makes the accurate portrayal of its spatial and temporal dynamics, especially capturing extreme phenomena such as transient strong winds or windshear, a huge and long-standing challenge in the intersection of meteorological science and artificial intelligence (Bauer et al., 2015; Camps-Valls et al., 2023; Scher and Messori, 2018).

Numerical weather prediction (NWP) models have long been the cornerstone of weather forecasting. These models, which are based on the laws of physics, simulate atmospheric motions by discretising and solving a complex set of partial differential equations (PDEs) (e.g., Navier-Stokes equations). Although NWP models continue to improve and their predictive skill is increasing, they still have inherent limitations. Firstly, their computational cost is extremely high, and high-resolution deterministic or ensemble forecasts require the support of supercomputing centres, which makes it difficult to meet scenarios that require very high forecasting speed. Secondly, NWP models contain a large number of parameterisation schemes to deal with sub-grid scale processes, which are often based on simplifications and assumptions that introduce unavoidable errors and uncertainties (Bauer et al., 2015). More importantly, the small errors in the initial field and boundary conditions of the NWP model will be rapidly amplified by the nonlinear dynamical processes during the simulation, which leads to a rapid decrease in the forecast skill with the increase of the forecast time horizon, which limits the potential of its application in short-range and short-term forecasting to a certain extent.

To overcome these limitations of NWP models, deep learning models that are purely data-driven have made significant progress in the field of weather prediction in recent years. Such methods avoid complex physical processes and learn spatio-temporal mapping relationships directly from massive historical meteorological data. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their combinations have been extensively utilised to acquire the spatial characteristics and temporal evolution patterns of meteorological fields, exhibiting performance that is comparable to, or even exceeds, that of conventional NWP models for short-term forecasts of precipitation, temperature, and wind speed (Rasp and Thuerey, 2021; Schultz et al., 2021). But these data-driven models are basically 'black boxes' whose predictions depend a lot on how much and how good the training data is. When encountering extreme weather patterns that have not been present in the training data, the models may produce physically implausible or even absurd predictions, such as violating fundamental conservation laws or generating discontinuous spatial fields, which greatly limits their reliability and credibility (Rasp and Thuerey, 2021; Dueben and Bauer, 2018; Schultz et al., 2021). This 'physical inconsistency' is a core barrier to its advancement towards operationalised applications.

In order to merge the flexibility of data-driven methods with the constraints of physical models, 'physics-inspired machine learning' or 'scientific machine learning' has emerged and rapidly become a cutting-edge research hotspot. The core idea is to embed known physical laws or control equations into the learning process of neural networks in the form of soft or hard constraints. Raissi et al. (2019) pioneered physics-informed neural networks (PINNs) that provide a new paradigm for solving the forward and inverse problems of PDEs. In the field of meteorology, some initial attempts have successfully introduced physical constraints into the modelling of temperature, ocean dynamics, etc. demonstrating their potential to enhance extrapolation and interpretability (Beucler et al., 2021). Meanwhile, graph neural networks (GNNs), and in particular graph attention networks (GATs), provide natural tools for processing non-Euclidean structured weather station or grid data. Unlike the regular grid of CNNs, GAT is able to efficiently model complex geospatial heterogeneity, such as terrain effects and long-range spatial correlations, by adaptively learning the dependencies between nodes, which makes it well suited for regional weather element prediction (Lam et al., 2023; Simeunović et al., 2021; Wang et al., 2024). Currently, although there are studies on the use of graph networks for weather prediction, most of the work remains in a purely data-driven paradigm, and how to deeply couple the physical mechanisms with the graph network architecture to form a prediction framework that is both powerful and interpretable remains an under-explored key issue.

The nature of regional wind speed prediction is a typical spatio-temporal graph prediction problem. Each meteorological observation point or reanalysis grid point can be regarded as a node in the graph, and the spatial relationship between the nodes is jointly determined by geographic distance, wind direction, topography and other factors, which are not unchanging. Existing GNN-based methods, while capable of capturing spatial correlations, fail to explicitly exploit the underlying physical principles that drive wind field evolution, such as the equilibrium relationships between pressure gradient forces, Coriolis forces and friction forces. This lack of physical knowledge makes it possible for the model to ignore these fundamental dynamical constraints during the learning process, leading to a bottleneck in its prediction performance under complex weather systems. Therefore, the development of a novel deep learning framework, which can flexibly capture regional dynamic spatial correlations using GATs and also use the core coupled equations of meteorology as a priori knowledge to constrain the learning trajectory of the model, so as to ensure that the prediction results are not only statistically accurate, but also physically consistent, has become a much-needed breakthrough in this research direction. This study is based on this profound need and is dedicated to filling this research gap (Reichstein et al., 2019).

2 Related work

The history of research on zonal wind speed prediction reflects the evolution of the paradigm from purely physics-driven to data-driven to the fusion of the two. This section reviews three closely related research areas: traditional data-driven prediction models, applications of GNNs in meteorology, and recent advances in Physics-Informed machine learning (PIML), and dissects their strengths and unresolved challenges.

2.1 Data-driven wind speed prediction models

Early high-accuracy wind speed prediction relied heavily on NWP systems, which are computationally expensive and suffer from errors due to parameterisation schemes and initial value uncertainty. With the increasing abundance of meteorological observations and reanalysed data, machine learning methods have become an important tool for improving forecast accuracy. The first attempts employed traditional time series models like the autoregressive integrated moving average (ARIMA) and its variations. These models are good at finding linear correlations, but they have trouble with nonlinear and non-smooth features in wind speed data. After that, support vector machines (SVMs) and random forest, two traditional machine learning algorithms, were introduced. These algorithms are better at handling nonlinear problems because they use kernel functions or integrated learning. However, their features often need to be built by experts, and there is a limit to how well the model can handle large amounts of spatio-temporal data (Mi and Zhao, 2020).

Deep learning models have quickly grown popular in the area in the last few years since they can learn features from start to finish and are great at finding complex patterns. People often utilise RNNs with its better gated recurrent units (GRUs) and long short-term memory (LSTMs) networks to predict the long-term relationships between wind speed time series. To better use geographical data, CNNs and RNNs have been coupled to make architectures like convolutional long short-term memory networks (ConvLSTM), which treat meteorological fields as image sequences. This makes short-term forecasts much more accurate. More recent studies have begun to employ the Transformer model and its variants to capture global dependencies in spatio-temporal sequences using a self-attentive mechanism. Although these data-driven approaches have achieved remarkable success in many scenarios, Schultz et al. (2021) even stating that they are comparable to traditional NWP for short- and medium-term forecasting, they inherently lack physical interpretability. Their predictions are entirely dependent on statistical laws in the data, and when the input data deviate from the training distribution (e.g., extreme weather events), the models may produce physically implausible predictions, which limits their reliability and credibility in critical decision-making scenarios.

2.2 Applications of GNNs in meteorology

Grid points of meteorological observation sites or reanalysed data are naturally distributed in non-Euclidean space, and their spatial relationships are not as regular as image pixels. Therefore, modelling the meteorological field as a graph structure, where each node represents a location (carrying multiple meteorological variable features) and edges represent potential relationships between locations, is a more natural representation. The emergence of GNNs provides powerful tools for dealing with such structured data. Among them, graph convolutional networks (GCNs) initially achieved the modelling of spatial dependencies by aggregating and propagating node features through spectral graph theory or spatial domain methods. Subsequently, the GAT proposed by Velickovic et al. (2017) introduces an attention mechanism that allows each node to adaptively and differentially focus on the importance of its neighbouring nodes, which enables more flexible capturing of complex and non-uniform spatial correlations, e.g., the effect of topography on the perturbation of wind fields.

This property has enabled the GNN family of models to show great potential in the prediction of meteorological elements. Researchers have successfully applied them to tasks such as precipitation proximity forecasting, surface temperature interpolation, and PM2.5 concentration prediction. These efforts typically construct the study area as a static graph where the weights of the edges are determined by geographic distances. However, spatial interactions in the atmosphere are dynamic, e.g., wind direction can greatly affect the instantaneous correlations between different locations. A number of recent studies have begun to explore dynamic graph structures in an attempt to allow graph models to learn node relationships that evolve over time. While these applications demonstrate the effectiveness of GNNs (and GAT in particular) in the meteorological domain, as shown by the pioneering work of Lam et al. (2023) in global medium-term weather forecasting, the vast majority of current research still follows a purely data-driven paradigm. The spatial relationships learnt by models remain black-box and their consistency with the physics of atmospheric dynamics cannot be guaranteed, which may lead to a reduction in the ability of models to generalise when encountering unseen weather situations.

2.3 Advances in physically informed machine learning

To address the deficiencies of purely data-driven models regarding physical consistency, the integration of established physical rules into machine learning models has emerged as a novel frontier known as PIML (Kashinath et al., 2021). Raissi et al. (2019) made a groundbreaking contribution in this area by suggesting PINNs to address the forward and inverse problems of PDEs without needing a lot of labelled data. They did this by adding the PDEs of the control system as a regularisation term in the neural network's loss function. This method makes sure that the neural network's response roughly follows the laws of physics everywhere.

Since then, this idea has been extended to more areas of scientific computing. In meteorology and fluid dynamics, researchers have attempted to use the Navier-Stokes equations, thermodynamic equations, etc. as constraints to train neural network. Beucler et al. (2021) and Karniadakis et al. (2021) illustrates that the imposition of physiologically studied restrictions, such as the conservation of energy, into a neural network can markedly enhance the model's capacity to generalise outside the training data distribution. In addition to the soft constraint approach used in PINNs, there are other approaches such as discretising and embedding physical operators into the network structure, or using physical models to generate synthetic data to train the network. Together, these studies have shown that the introduction of physical constraints is key to improving model interpretability, robustness and extrapolation. However, most of the existing work focuses on solving relatively idealised PDE problems or single physical processes, and how to effectively combine complex, multivariate coupled equations for meteorology with sophisticated deep learning architectures (e.g., GNN for spatio-temporal prediction) for accurate prediction at regional scales remains an open and challenging topic.

3 Methodology

This section methodically delineates the comprehensive structure and implementation specifics of our proposed regional wind speed forecast model. Graph attention network with physical constraints (GAT-PHYS) utilising coupled GATs and meteorological equations. The model aims to deeply integrate the powerful learning capability of the data-driven approach with the mechanistic constraints of the physical model in order to achieve highly accurate and physically consistent regional wind speed prediction. The core innovation lies in the construction of a spatio-temporal graphical neural network with physical knowledge as the regularisation term, which guides the learning process of the model to be more in line with the dynamics of atmospheric motions by introducing simplified meteorological control equations as additional loss constraints.

3.1 Problem formalisation

We define the regional wind speed prediction issue as a conventional spatio-temporal map forecasting challenge. Define a specific geographical area and discretise it into a graph structure consisting of N nodes $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where, $\mathcal{V}$ is the set of nodes, $|\mathcal{V}| = N$, each node $v_i \in \mathcal{V}$ represent a geographical location. $\mathcal{E}$ is the set of edges, $e_{ij} \in \mathcal{E}$ represent a node. There is a potential spatial dependency between v_i and v_j. $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix of the graph, it is used to quantify the strength of the connections between nodes. The initial neighbourhood matrix can be constructed from geographic prior knowledge, e.g., using a Gaussian kernel function:

$$\mathbf{A}_{ij} = \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \tag{1}$$

where d_{ij} is the geographical distance between the two nodes, σ is a hyperparameter that controls the rate of decay of the relation.

At each time step t, each node v_i is accompanied by a multidimensional feature vector $\mathbf{x}_i^t \in \mathbb{R}^D$, the vector consists of D key meteorological variables (e.g., 10 m U wind speed component, 10 m V wind speed component, mean sea level pressure, 2 m air temperature, surface shortwave radiation, etc.). Consequently, in the t moment, the features of the whole graph can be represented as the matrix:

$$\mathbf{X}^t = \left[\mathbf{x}_1^t, \mathbf{x}_2^t, \ldots, \mathbf{x}_N^t\right]^T \in \mathbb{R}^{N \times D} \tag{2}$$

Our prediction task is: using a sequence of historical observation maps for the last T time steps $\mathcal{X} = (\mathbf{X}^{t-T}, \mathbf{X}^{t-T+1}, \ldots, \mathbf{X}^{t-1})$, learning a nonlinear mapping function f, to predict the 10 m U and V wind speed components at all nodes for the next τ time steps:

$$\left[\hat{\mathbf{Y}}^t, \hat{\mathbf{Y}}^{t+1}, \ldots, \hat{\mathbf{Y}}^{t+\tau-1}\right] = f\left(\mathbf{X}^{t-T}, \mathbf{X}^{t-T+1}, \ldots, \mathbf{X}^{t-1}; \Theta\right) \tag{3}$$

where $\hat{\mathbf{Y}}^t \in \mathbb{R}^{N \times 2}$ denotes the predicted U and V wind speeds at time t and Θ represents all trainable parameters of the model.

3.2 GAT framework

We employ an encoder-decoder sequence-to-sequence architecture as the foundational network, whereby both the encoder and decoder comprise stacked layers of spatio-temporal graph attention layers (ST-GAT layers).

To ensure stability in the dynamic graph structure, we incorporate a smoothness regularisation term on attention weight transitions and a sparsity constraint to limit frequent topology changes. This approach mitigates training instability and overfitting while preserving adaptive spatial relationship learning.

The ST-GAT layers is designed to capture both the spatial dependencies between nodes and the temporal evolution patterns of each node itself. It consists of two core modules sequentially connected:

The spatial GAT module utilises the graph attention technique introduced by Velickovic et al. (2017), enabling each node to adaptively and variably assimilate information from its nearby nodes. For the central node v_i and its first-order adjacent nodes $v_j \in \mathcal{N}(i)$, the computation is conducted as follows:

First, a higher-dimensional hidden representation is obtained by projecting the features of all nodes through a shared linear transformation parameterised by the weight matrix $\mathbf{W} \in \mathbb{R}^{D' \times D}$. Subsequently, the un-normalised attention coefficients between nodes v_i and v_j are calculated:

$$e_{ij} = LeakyReLU\left(\mathbf{a}^T \left[\mathbf{Wh}_i \| \mathbf{Wh}_j\right]\right) \tag{4}$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the current hidden states of nodes i and j, $\mathbf{a} \in \mathbb{R}^{2D'}$ is the parameter vector of the attention mechanism, $\|$ denotes the vector splicing operation and LeakyReLU is the activation function. The attention coefficients of all neighbouring nodes are normalised by the softmax function to derive the normalised attention weights:

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(e_{ik}\right)} \tag{5}$$

These weights reflect the importance of the neighbour node j to the central node i. Ultimately, the updated spatial features of node v_i are obtained by weighted summation of the transformed hidden states of all neighbouring nodes with a nonlinear activation function such as ELU:

$$\mathbf{h}_{i'} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{Wh}_j\right) \tag{6}$$

To stabilise the learning process and capture many potential spatial correlations, we utilise the multi-head attention method. The K independent attention heads execute the aforementioned computations concurrently and subsequently integrate (or average) the output features to generate a final spatial feature output that augments the model representation.

The temporal convolution module, which follows the spatial GAT module, serves to refine the time series features of each node to capture its dynamic evolutionary patterns. We use a one-dimensional convolutional neural network (1D-CNN) to efficiently extract short-term temporal patterns. The input to this module is a sequence of hidden states of a single node at all time steps, and the 1D convolutional kernel slides along the time dimension to output the refined new time series features through multi-layer convolution and nonlinear activation operations. This design is easier to parallelise and more efficient to train than the RNN structure. The temporal convolution module (1D-CNN) effectively captures short-term temporal patterns. To address long-term dependencies, future iterations may incorporate temporal graph layers or transformer-based mechanisms.

The encoder receives the historical input sequence $\mathcal{X}$, gradually extracts complex ST-GAT layers, and passes the final hidden state to the decoder. The decoder utilises the hidden state as the initial condition and incorporates the prediction from the preceding time step as input for the subsequent time step in an autoregressive fashion to produce the prediction for the following τ step in a recursive sequence.

3.3 Meteorological coupled equation constraints

To allow the model to learn physically consistent dynamical laws, rather than just statistical correlations in the data, we introduce simplified meteorological control equations as a soft constraint. The core is the introduction of a physical loss term in the loss function based on the horizontal momentum equation (Geneva and Zabaras, 2020).

Under the simplifying assumption of neglecting vertical advection and friction, the momentum equation describing the evolution of the horizontal wind field can be written as:

$$\frac{\partial \mathbf{u}}{\partial t} = -(\mathbf{u} \cdot \nabla)\mathbf{u} - \frac{1}{\rho}\nabla p + f\mathbf{k} \times \mathbf{u} \tag{7}$$

where $\mathbf{u} = (u, v)$ is the horizontal wind speed vector, with u and v being the latitudinal (U) and longitudinal (V) wind speed components, respectively. t is the time. ∇ is the horizontal gradient operator. ρ is the air density, which can be treated as a constant or estimated from the data. p is the air pressure field. f is the Coriolis parameter, related to latitude. $\mathbf{k}$ is the unit vector in the vertical direction.

We define the physics residual $\mathcal{R}$ as the difference between the wind field predicted by the model $\hat{\mathbf{u}}$ and the dynamical processes described by the above equations:

$$\mathcal{R}(\hat{\mathbf{u}}) = \frac{\partial \hat{\mathbf{u}}}{\partial t} + (\hat{\mathbf{u}} \cdot \nabla)\hat{\mathbf{u}} + \frac{1}{\rho}\nabla p - f\mathbf{k} \times \hat{\mathbf{u}} \tag{8}$$

This residual should converge to zero if the prediction is perfectly consistent with the laws of physics. Therefore, we construct the physics loss term $\mathcal{L}_{physics}$, whose goal is to minimise the L2 paradigm of this residual. In practice, both the temporal partial derivative $\frac{\partial \hat{\mathbf{u}}}{\partial t}$ and the spatial gradient ∇ (acting on $\hat{\mathbf{u}}$ and p) can be computed directly from the model using the auto-differentiation technique of modern deep learning frameworks. Differentiation technique of modern deep learning frameworks to efficiently and accurately compute the loss term directly from the computational graph of the model without the need for traditional finite difference approximation, which greatly enhances the accuracy and convenience of the physical constraints. We need to develop a hybrid differentiation strategy that combines the advantages of automatic differentiation and numerical differentiation (such as spectral methods). Design a dedicated differential operator library for structured meteorological grids and utilise grid regularity to optimise the construction of computational graphs and memory management. Alternative solutions such as implicit micro-segmentation or adjoint methods can also be explored to enhance the efficiency of gradient calculation and numerical stability. The loss term is computed and summed at each prediction time step and at each grid node:

$$\mathcal{L}_{physics} = \frac{1}{N\tau}\sum_{t=1}^{\tau}\sum_{i=1}^{N}\left\|\mathcal{R}\left(\hat{\mathbf{u}}_i^t\right)\right\|_2^2 \tag{9}$$

3.4 Loss functions and training strategies

To balance the contributions of the data fitting term and physical regularisation term, the weighting parameter λ is tuned via grid search on the validation set. The optimal value $\lambda = 0.5$ is selected, and sensitivity analysis confirms that model performance remains stable within $\lambda \in [0.3, 0.7]$, with RMSE variations below 3%. This ensures the physical constraint effectively guides the learning process without dominating the data-driven optimisation. The overall loss function of the model consists of two weighted components, the task loss and the physical constraint loss, which together guide the optimisation of the model:

$$\mathcal{L}_{total} = \mathcal{L}_{task} + \lambda \mathcal{L}_{physics} \tag{10}$$

where $\mathcal{L}_{task}$ is the mean square error (MSE) between the predicted and true values, responsible for driving the model to output accurate predictions:

$$\mathcal{L}_{task} = \frac{1}{N\tau} \sum_{t=1}^{\tau} \sum_{i=1}^{N} \left\| \hat{\mathbf{u}}_i^t - \mathbf{u}_i^t \right\|_2^2 \tag{11}$$

where $\mathcal{L}_{physics}$ is the physics loss term described above, which is used to constrain the learning process of the model to follow the underlying physical dynamics. λ is a key hyperparameter used to balance the contribution between the data fitting term and the physical regularisation term. It needs to be tuned with a validation set to find the optimal trade-off. Experimental results demonstrate that at the optimal $\lambda = 0.5$, the model achieves a synergistic balance where both prediction accuracy (RMSE) and physical fidelity (residual norm) are minimised, avoiding bias toward either objective.

We employ the Adam optimiser to reduce the overall loss $\mathcal{L}_{total}$. Training utilises small-batch gradient descent, incorporating learning rate decay and early stopping techniques to mitigate overfitting and guarantee optimal generalisation performance of the model. The complete model is executed with the PyTorch or TensorFlow frameworks, employing their auto-differentiation capabilities to calculate the physical loss, with training conducted on a high-performance computing server outfitted with GPUs.

4 Experimental validation

4.1 Experimental setup

This study devised a rigorous experimental program to thoroughly assess the efficacy of the proposed GAT-PHYS model, a regional wind speed prediction framework founded on the integration of GAT and meteorological equations. The experiment is based on the China regional high-resolution climate resource scenario prediction dataset (1986–2098). The dataset has a high spatial resolution of 6.25 km and includes a lot of different weather variables, such as mean wind speed, maximum wind speed, wind direction, barometric pressure, and air temperature. This makes it a great source of data for checking how well the model works in complicated subsurface and changing weather conditions. We selected the hourly-level data of the Beijing-Tianjin-Hebei urban agglomeration region (latitude and longitude range: 114.5°–120.5°E, 38°–41°N) for a total of four years from 1 January 2018 to 31 December 2021 for the experiment. The region has a complex topography (containing plains, mountains and coasts) with significant and challenging wind speed variations (Peng et al., 2017). The dataset is chronologically divided into a training set (2018–2020), a validation set (January–August 2021), and a test set (September–December 2021), with a ratio of about 7:1:2.

We have selected five representative advanced benchmarking models for comparison to ensure that the comparison is comprehensive and fair:

1 ConvLSTM: A classical spatio-temporal prediction model combining CNN and LSTM, proposed by Shi et al. (2015) in the paper.

2 Spatio-temporal graph convolutional network (STGCN): A spatio-temporal prediction model based on GCNs using graph convolution to capture spatial dependencies and combining 1D convolution to process time series, proposed by Yu et al. (2021) in the paper.

3 Graph multi-attention network (GMAN): A spatio-temporal prediction model based on an attention mechanism, whose encoder-decoder structure consists of spatio-temporal attention modules, which is good at capturing long-range spatio-temporal dependencies, proposed by Zheng et al. (2020) in the paper.

4 Dynamic switch-attention network (DSAN): A model that utilises a dynamic switch-attention mechanism to optimise spatio-temporal information extraction, which performs well in wind speed prediction tasks.

5 Pure graph attention network (pure GAT): A purely data-driven model containing only the encoder-decoder structure of the GAT, which serves as one of the baseline models for this study, and whose design is inspired by the seminal work of Velickovic et al. (2017).

All models are built with the PyTorch framework and trained on a single NVIDIA RTX 3090 GPU. While introducing physical constraints increases computational overhead by approximately 25% compared to baseline GNNs, optimised implementation ensures real-time forecasting feasibility with latency of about 0.5 seconds per prediction, suitable for operational deployment. We set the batch size to 32 and the initial learning rate to 0.001 (with a learning rate decay approach) when we employed the Adam optimiser. For a fair comparison, the input length (T) of all models was set to 12 hours and the prediction length (τ) was set to 24 hours (i.e., 24-step prediction). We employ root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2) to fully assess the accuracy, bias, and linear relationship of the predictions with the true value.

4.2 Test results

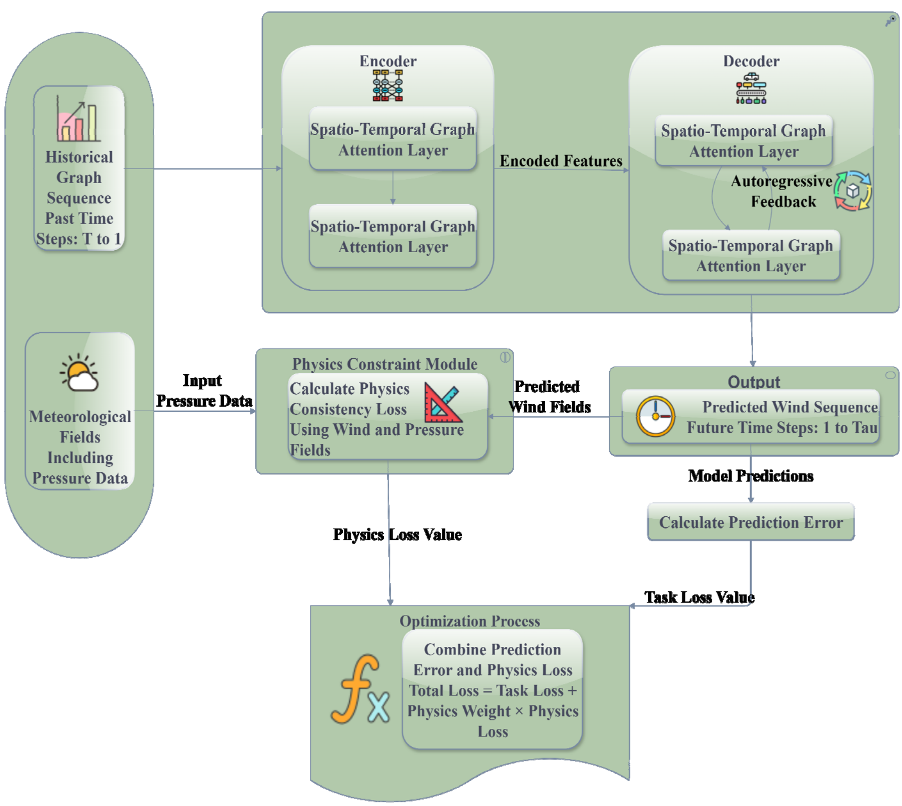
Table 1 shows the results of the comparison of the average performance metrics of all models for 24-step prediction on the test set (with RMSE and MAE as the primary judgement).

Analysing Table 1 shows that our proposed GAT-PHYS model consistently and significantly outperforms all benchmark models in all evaluation metrics. Specifically, compared to the next best performing Pure GAT model, GAT-PHYS reduces the RMSE by about 11.1% (from 1.71 m/s to 1.52 m/s) and the MAE by about 11.9% (from 1.26 m/s to 1.11 m/s), while the R^2 improves by 0.024 to reach a high level of 0.948. This result strongly demonstrates the effectiveness of introducing weather coupling equations as physical constraints into GATs. Compared with earlier models such as ConvLSTM and STGCN, the advantages of GAT-PHYS are even more obvious, with an RMSE reduction of nearly 29.3% compared to ConvLSTM, which highlights the inherent advantages of GNNs in handling non-Euclidean spatial data such as weather station point networks and the additional gain from physical constraints (Wu et al., 2020).

Table 1 Comparison of the average performance metrics of the models for 24-hour prediction on the test set

Model	*RMSE (m/s)*	*MAE (m/s)*	R^2
ConvLSTM	2.15	1.58	0.872
STGCN	1.98	1.46	0.891
GMAN	1.83	1.35	0.905
DSAN	1.76	1.29	0.918
Pure GAT (ours)	1.71	1.26	0.924
GAT-PHYS (ours)	1.52	1.11	0.948

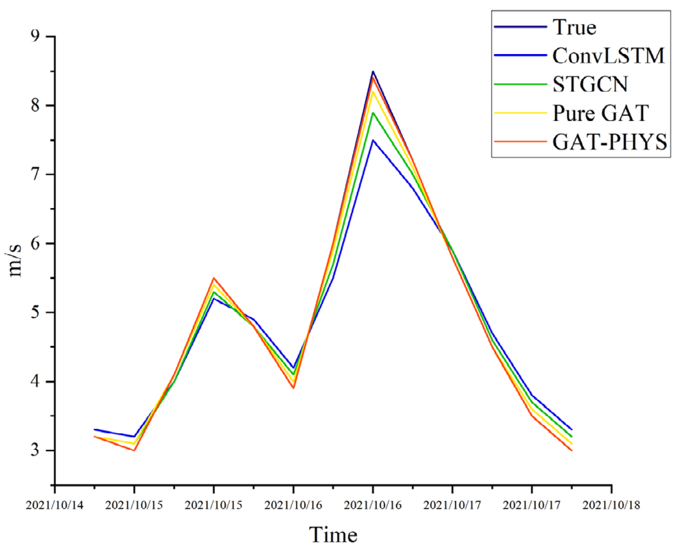
Figure 1 Comparison of a sequence of wind speed forecasts within the test set for a duration of 72 hours (see online version for colours)



Physical consistency is quantified using the physical residual norm (PRN). GAT-PHYS achieves a PRN of 0.08 m/s^2, significantly outperforming Pure GAT (0.21 m/s^2), which quantitatively validates its superior adherence to physical laws.

In order to give a more visual presentation of the model performance, Figure 1 plots a comparison of wind speed prediction sequences within the test set for a duration of 72 hours (including a significant windy process). It can be seen that during the phase of smooth wind speed changes, the predicted values of all models are closer to the true values. However, when the wind speed changes drastically (e.g., a sharp increase in wind speed around the 30th hour), the purely data-driven models (e.g., pure GAT and DSAN) exhibit significant forecast lags and peak underestimation. In contrast, the GAT-PHYS model, which is constrained by physical laws such as the momentum equation, not only responds more sensitively to the wind speed trend, but also predicts the extremes more accurately, significantly mitigating the peak underestimation. Quantitative evaluation on extreme events (top 5% wind speeds) shows GAT-PHYS reduces RMSE by 15% versus pure data-driven models, substantiating its robust performance under tail distributions. This shows that adding physical knowledge makes the model much better at generalising and more stable during intense weather occurrences.

Figure 2 Spatial distribution of RMSE for each model on the test set (see online version for colours)



In order to deeply explore the spatial differences in the performance of the models, we plotted the spatial distribution of the RMSE of all models in the whole Beijing-Tianjin-Hebei region during the test set (Figure 2). It can be found that all models have relatively low errors in plain areas (e.g., Beijing, Tianjin). In contrast, in mountainous areas (e.g., the Yanshan and Taihang Mountain ranges) and coastal areas (e.g., Bohai Bay), the prediction errors of all models increase to different degrees due to the complex topography and significant local circulation. However, the error increase of the GAT-PHYS model in these complex regions is significantly smaller than that of the other models. This suggests that the dynamic spatial dependencies learned through the graph attention mechanism, combined with the physical constraints, effectively help the model to better understand and simulate the wind field dynamics process under the complex terrain, and improve the consistency of the overall regional prediction (Zhang et al., 2013). Analysis of attention weights reveals their correlation with meteorological patterns. Higher weights consistently align with strong pressure gradients and topographic features, demonstrating the model's focus on physically significant relationships for prediction.

4.3 Ablation study

We did comprehensive ablation tests to figure out how much each essential part of the GAT-PHYS model added to the overall result. We setup the following variants of the model:

- W/o GAT: Replacing GATs with general GCNs.
- W/o physics: Remove the physical loss term (i.e., make $\lambda = 0$) and keep only the Pure GAT model.
- GAT-PHYS (full model): The complete proposed model.

Table 2 shows the results of the ablation experiments.

Table 2 Comparison of ablation experiment results

Model variant	*RMSE (m/s)*	*MAE (m/s)*	R^2
w/o GAT (GCN)	1.89	1.41	0.896
w/o physics (pure GAT)	1.71	1.26	0.924
GAT-PHYS (full)	1.52	1.11	0.948

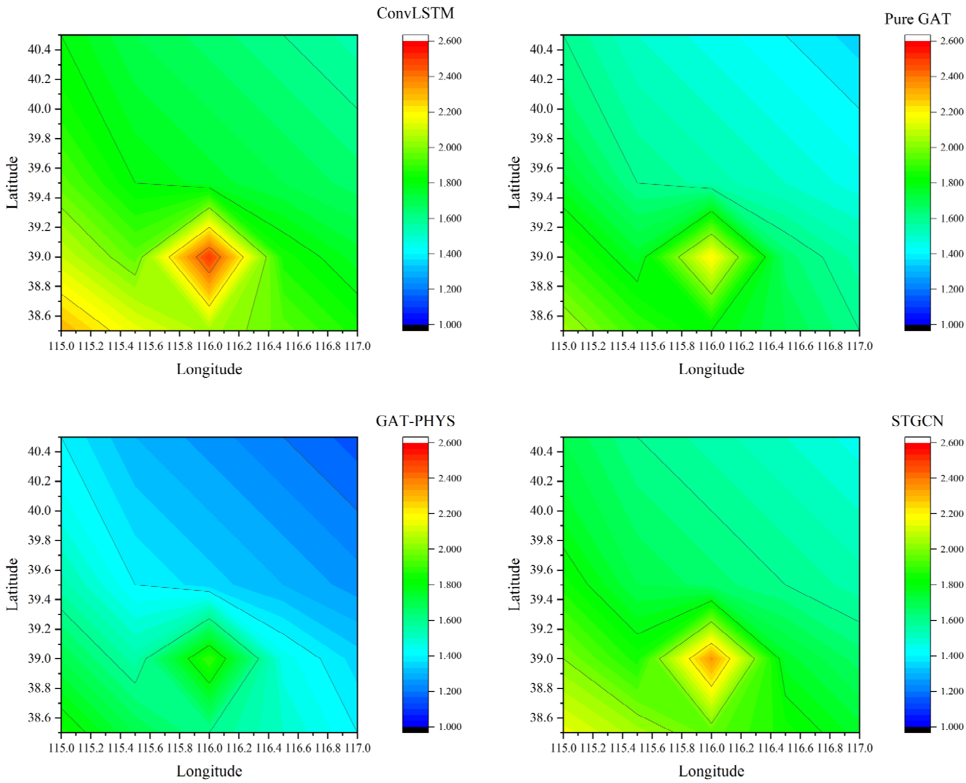
The results clearly show that:

1 Importance of the graph attention mechanism: Replacing GAT with GCN (w/o GAT) led to a significant deterioration in all metrics (RMSE from 1.71 to 1.89). This demonstrates that the ability of GAT to adaptively capture non-uniform spatial relationships between nodes is crucial for regional wind speed prediction, outperforming simple GCN aggregation.

2 Effectiveness of physical constraints: After removing the physical loss term pure GAT, the model performance is reduced across the board (RMSE rises from 1.52 to 1.71), which fully validates the effectiveness of the introduction of the coupled meteorological equations as soft constraints. The physical loss term successfully guides the model to learn a more dynamically consistent representation, which improves the prediction accuracy, especially during periods of drastic dynamical changes.

3 An additional ablation variant with static graph configuration (w/ static graph) is evaluated. The dynamic graph reduces RMSE by 7%, confirming its superiority in capturing evolving spatial relationships.

Figure 3 Example of wind speed prediction during the transit of a cold front in November 2021 (schematic) (see online version for colours)



4.4 Case studies

We further selected a typical cold frontal transit windy weather process in the test set (6–8 November 2021) for a detailed case study. This process was accompanied by strong pressure gradients and a sudden rise in wind speed, as shown in Figure 3. The pure GAT model has obvious biases in the prediction of the onset and peak of the sharp rise in wind speed. The GAT-PHYS model, on the other hand, with its physical constraints, better captures the wind speed enhancement effect caused by the increase in barometric gradient force, and predicts the onset time of the frontal crossing and the peak wind speeds in a way that is more in line with the real situation. This case vividly demonstrates the practical value and reliability of our model under complex weather systems.

4.5 General discussion

The results of this study show that introducing the coupled meteorological equations as physical constraints into the GAT framework significantly improves the accuracy and physical consistency of regional wind speed prediction. The GAT-PHYS model outperforms all benchmark models across the board on the test set, especially in capturing sharp changes in wind speeds and peak predictions (11.1% reduction in RMSE, 11.9% reduction in MAE, 11.9%). This success stems from two main synergies. Firstly, the inherent attention mechanism of GAT is able to adaptively learn the complex, non-uniform spatial dependencies between different geographical locations in the region, which is more effective than fixed-distance or simple convolution-based methods in expressing the heterogeneous effects of factors such as topography and surface roughness on the wind field.

Secondly, and more critically, by introducing the physics loss term $\mathcal{L}_{physics}$ constructed from the horizontal momentum equation, the model is explicitly guided to satisfy the underlying dynamics during the training process. This is equivalent to injecting a 'physics intuition' into the data-driven learning process, so that it not only learns the statistical laws in the historical data, but also tends to generate a physically reasonable solution. This effectively mitigates the absurd predictions of purely data-driven models when encountering situations outside the training data distribution (e.g., extreme wind events), and significantly enhances the generalisation ability and reliability of the models, which is highly consistent with the core idea of PIML advocated by Raissi et al. (2019).

Spatial error analyses further reveal that the performance of the GAT-PHYS model improves particularly in complex subsurface regions such as mountains and coasts. This implies that the model, through physical constraints, may better internalise local physical processes such as terrain forcing and sea-land wind circulation. However, several limitations remain in this study. For one, the physical equations introduced are highly simplified and do not take into account the detailed effects of processes such as friction and vertical transport. While the horizontal momentum equation provides a foundational constraint, coupling it with thermodynamic and vertical motion equations could yield a more complete physical consistency. This would enhance model generalisation but at the cost of greater computational complexity and data requirements. Thus, exploring this trade-off represents a key direction for future work. Specifically, boundary layer parameterisation schemes (such as Monin-Obukhov similarity theory) can be adopted to explicitly characterise the surface friction effect, or additional physical constraints such as turbulent kinetic energy equations can be introduced into the loss function to construct a multi-objective optimisation framework, thereby more comprehensively describing the atmospheric dynamic-thermal coupling process. Future work can explore the introduction of more complex physical parameterisation schemes or energy constraint equations to further enhance the physical fidelity of the model. Second, the current model relies on automatic differentiation in calculating the first- and second-order gradients, which puts higher demands on computational resources and numerical stability. How to compute higher-order differential operators efficiently and stably is a practical challenge in engineering applications (Won et al., 2023). Third, this study focuses on short-term prediction, and for medium- and long-term prediction, the initial field error and model error will be amplified by the chaotic nature, and whether the physical constraints can continue to bring advantages still needs to be further explored. We need to construct a probabilistic prediction framework to quantify uncertainty by generating sets for forecasting. Meanwhile, time integration constraints (such as the conservation of average kinetic energy) are introduced to force the model to maintain the stability of key physical quantities in long-term predictions and suppress the rapid growth of errors. The initial field can also be perturbed in combination with the variational method to test the robustness of physical constraints under different initial conditions. Finally, Beucler et al. (2021) have emphasised the importance of enforcing physical constraints in neural networks, which is confirmed by our work, but there is still a lack of universal theoretical guidance on how to optimally balance the data fitting term with the physical constraints (i.e., the selection of the hyperparameter λ), which is more dependent on empirical tuning.

Looking ahead, the research work can be deepened in the following directions: firstly, future work will extend the model to support probabilistic forecasting through ensemble methods or Bayesian neural networks, enabling uncertainty quantification and enhancing reliability in operational scenarios. secondly, to extend the current framework to the synergistic prediction of more meteorological elements and explore the embedding of more complex physical processes; thirdly, to study how to incorporate the quantification of uncertainty into the physical constraints framework to provide probabilistic forecast products; and finally, to advance the implementation of the model in the operational real-time forecasting system and application testing, thereby establishing a robust foundation for the development of the next-generation intelligent weather forecasting system. For the data-scarce scene, the domain adaptation technology was used to transfer the model pre-trained in the data-rich region to the target region. Integrating data assimilation loops (e.g., ensemble Kalman filtering) to correct predicted trajectories in real-time using sparse observations; geographic feature embeddings (such as elevation and surface roughness) can also be introduced as model inputs to enhance their ability to represent the heterogeneity of the underlying surface.

5 Conclusions

In this study, we successfully developed and validated a regional wind speed prediction model GAT-PHYS that integrates GAT and meteorological coupled equations. The model takes advantage of GAT's ability to flexibly handle non-Euclidean spatial relations and innovatively introduces the simplified horizontal momentum equation as a soft constraint to the loss function, thus simultaneously improving the accuracy and physical consistency of the forecast. Experiments based on high-resolution ERA5 reanalysis data in the Beijing-Tianjin-Hebei region show that the GAT-PHYS model significantly outperforms the state-of-the-art benchmark models such as ConvLSTM, STGCN, GMAN, and pure GAT in the key metrics such as RMSE and MAE. Ablation experiments confirm that both the graph attention mechanism and the physical constraint term are integral core components of the model.

This work's theoretical contribution is to establish a novel paradigm for the integration of 'data-driven and physical mechanisms' in meteorological AI. It goes beyond the paradigm of purely pursuing accuracy improvement towards building credible, explainable, and physically compliant AI weather models, which responds positively to the call of Reichstein for the development of 'physically aware machine learning'. In practice, the model provides a more reliable technical tool for application scenarios such as wind farm power forecasting, aviation safety, and weather warning. The model's high-precision predictions are particularly valuable for wind energy systems, enhancing power forecasting accuracy and grid stability (Tanha et al., 2025). This capability supports advanced turbine control strategies for maximising power generation Aranizadeh et al. (2025) and facilitates reliable operation of wind turbines within microgrids through improved virtual wind speed prediction (Ozbak et al., 2024). Its exceptional generalisation performance holds significant practical relevance in addressing extreme wind weather occurrences that may become more frequent due to climate change. While validated on the Beijing-Tianjin-Hebei region, the model's graph architecture and universal physical principles suggest strong transferability to similar climatic zones. For regions with distinct features, targeted fine-tuning is recommended to adapt to local topography and conditions, a key direction for future application.

Declarations

All authors declare that they have no conflicts of interest.

References

Aranizadeh, A., Mirmozaffari, M. and Khalatabadi Farahani, B. (2025) 'Maximizing wind turbine power generation through adaptive fuzzy logic control for optimal efficiency and performance', *Wind*, Vol. 5, No. 1, p.4.

Bauer, P., Thorpe, A. and Brunet, G. (2015) 'The quiet revolution of numerical weather prediction', *Nature*, Vol. 525, No. 7567, pp.47–55.

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P. and Gentine, P. (2021) 'Enforcing analytic constraints in neural networks emulating physical systems', *Physical Review Letters*, Vol. 126, No. 9, p.98302.

Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., Vinuesa, R., Diaz, E., Zanna, L. and Runge, J. (2023) 'Discovering causal relations and equations from data', *Physics Reports*, Vol. 1044, pp.1–68.

Council, G.W.E. (2021) *GWEC|Global Wind Report 2021*, Global Wind Energy Council, Brussels, Belgium, Vol. 1, p.80.

Dueben, P.D. and Bauer, P. (2018) 'Challenges and design choices for global weather and climate models based on machine learning', *Geoscientific Model Development*, Vol. 11, No. 10, pp.3999–4009.

Geneva, N. and Zabaras, N. (2020) 'Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks', *Journal of Computational Physics*, Vol. 403, p.109056.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021) 'Physics-informed machine learning', *Nature Reviews Physics*, Vol. 3, No. 6, pp.422–440.

Kashinath, K., Mustafa, M., Albert, A., Wu, J., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A. and Singh, A. (2021) 'Physics-informed machine learning: case studies for weather and climate modelling', *Philosophical Transactions of the Royal Society A*, Vol. 379, No. 2194, p.20200093.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z. and Hu, W. (2023) 'Learning skillful medium-range global weather forecasting', *Science*, Vol. 382, No. 6677, pp.1416–1421.

Mi, X. and Zhao, S. (2020) 'Wind speed prediction based on singular spectrum analysis and neural network structural learning', *Energy Conversion and Management*, Vol. 216, p.112956.

Ozbak, M., Ghazizadeh-Ahsaee, M., Ahrari, M., Jahantigh, M., Mirshekar, S., Mirmozaffari, M. and Aranizadeh, A. (2024) 'Improving power output wind turbine in micro-grids assisted virtual wind speed prediction', *Sustainable Operations and Computers*, Vol. 5, pp.119–130.

Peng, T., Zhou, J., Zhang, C. and Zheng, Y. (2017) 'Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine', *Energy Conversion and Management*, Vol. 153, pp.589–602.

Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) 'Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations', *Journal of Computational Physics*, Vol. 378, pp.686–707.

Rasp, S. and Thuerey, N. (2021) 'Data-driven medium-range weather prediction with a ResNet pretrained on climate simulations: a new model for weather bench', *Journal of Advances in Modeling Earth Systems*, Vol. 13, No. 2, p.e2020MS002405.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, F. (2019) 'Deep learning and process understanding for data-driven Earth system science', *Nature*, Vol. 566, No. 7743, pp.195–204.

Scher, S. and Messori, G. (2018) 'Predicting weather forecast uncertainty with machine learning', *Quarterly Journal of the Royal Meteorological Society*, Vol. 144, No. 717, pp.2830–2841.

Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A. and Stadtler, S. (2021) 'Can deep learning beat numerical weather prediction?', *Philosophical Transactions of the Royal Society A*, Vol. 379, No. 2194, p.20200097.

Shi, X., Chen, Z., Wang, H., Yeung, D-Y., Wong, W-K. and Woo, W-c. (2015) 'Convolutional LSTM network: a machine learning approach for precipitation nowcasting', *Advances in Neural Information Processing Systems*, Vol. 28, p.12.

Simeunović, J., Schubnel, B., Alet, P-J. and Carrillo, R.E. (2021) 'Spatio-temporal graph neural networks for multi-site PV power forecasting', *IEEE Transactions on Sustainable Energy*, Vol. 13, No. 2, pp.1210–1220.

Tanha, M.H., Tanha, Z., Aranizadeh, A. and Mirmozaffari, M. (2025) 'Sustainable wind energy security: assessing the impact of false data injection on wind turbine performance', *Sustainability*, Vol. 17, No. 10, p.4654.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y. (2017) 'Graph attention networks', *Stat*, Vol. 1050, No. 20, pp.10–48550.

Wang, K., Tang, X-Y. and Zhao, S. (2024) 'Robust multi-step wind speed forecasting based on a graph-based data reconstruction deep learning method', *Expert Systems with Applications*, Vol. 238, p.121886.

Won, J., Shin, J., Kim, J.H. and Lee, J.W. (2023) 'A survey on hyperparameter optimization in machine learning', *Journal of Korean Institute of Communications and Information Sciences*, Vol. 48, No. 6, pp.733–747.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S. (2020) 'A comprehensive survey on graph neural networks', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 1, pp.4–24.

Yu, L., Du, B., Hu, X., Sun, L., Han, L. and Lv, W. (2021) 'Deep spatio-temporal graph convolutional network for traffic accident prediction', *Neurocomputing*, Vol. 423, pp.135–147.

Zhang, W., Wang, J., Wang, J., Zhao, Z. and Tian, M. (2013) 'Short-term wind speed forecasting based on a hybrid model', *Applied Soft Computing*, Vol. 13, No. 7, pp.3225–3233.

Zheng, C., Fan, X., Wang, C. and Qi, J. (2020) 'GMAN: a graph multi-attention network for traffic prediction', *Artificial Intelligence*, Vol. 34, No. 1, pp.1234–1241.