Wind Speed Forecasting in Kazakhstan using Deep Learning

by

Zholaman Kuangaliyev

Submitted to the Department of Robotics and Mechatronics
in partial fulfillment of the requirements for the degree of

Master of Science in Robotics

at the

NAZARBAYEV UNIVERSITY

Apr 2025

Author . . . .
Department of Robotics and Mechatronics
Apr 29, 2025

Certified by . . . .
Ton Duc Do
Associate Professor
Thesis Supervisor

Accepted by . . . .
Vassilios D. Tourassis
Dean, School of Engineering and Digital Sciences

Wind Speed Forecasting in Kazakhstan using Deep Learning

by

Zholaman Kuangaliyev

Submitted to the Department of Robotics and Mechatronics
on Apr 29, 2025, in partial fulfillment of the
requirements for the degree of
Master of Science in Robotics

Abstract

Accurate wind speed forecasting is crucial for grid stability and for renewable energy sources optimization. In this study we will be evaluating performance of 5 models: 1D Convolutional Neural Network (1D CNN), Long-Short Term Memory (LSTM), WaveNet, Transformer, and Swin Transformer for Time Series (Swin4TS); which was used/adapted for wind speed forecasting across different regions of the Kazakhstan. We utilized 3 different datasets: NASA, KazHydroMet, and one we collected ourselves using the Yurt facility at the territory of Nazarbayev University. Models' evaluation is set across different forecast horizons: ultra-short-term (ULTF), short-term (STF), medium-term (MTF), and long-term (LTF). Our results showed that there is no model that universally superior for every dataset and every forecast horizon. LSTM showed excellent performance for USTF oncomplex and cyclic data. WaveNet showed an ability to adapt for long-term forecasting patterns in the challenging NASA dataset. In the KazHydroMet dataset's domain the Transformer model was proven to be the top performer. 1D CNN showed a baseline consisted mid-tier performance across all datasets. The novel architecture Swin4TS in general showed poor results, but managed to narrow the gap in LFT and was fairly comparable with other models.

Thesis Supervisor: Ton Duc Do
Title: Associate Professor

Acknowledgments

I want to express my sincere gratitude to my supervisor Professor Ton Duc Do and co-supervisor Professor Anh Tu Nguyen, for their support and guidance throughout my master thesis.

My special gratitude goes to my family and my wife - Katarina Kuangaliyeva, you believed in me more then I believed in myself.

I'm extremely grateful to my friend - Mukhamejan Talap, whom I met during my masters. Together we succeeded in many hackathons and I believe there is more in the future.

Lastly, I would like to mention, KazHydroMet, for providing Kazakhstan's wind speed data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the modern world, sustainable and renewable energy sources play an important role as a result of the depletion of fossil fuels. Many countries aim to reduce fossil fuel consumption and minimize greenhouse emission. Wind energy brings clean and sustainable energy. However, integrating wind farms efficiently into existing power grid systems is dependent on the ability to accurately forecast wind patterns based on historical data. Precise and low error forecasting will allow better energy production scheduling, enhance grid stability, reduce operational costs, and facilitate energy trading.

1.1 Motivation for Wind Speed Forecasting in Kazakhstan

Kazakhstan's energy possess vast steppes and open plains, which brings a great potential for wind energy, particularly in the Dzungarian Gates and along the coast of the Caspian Sea [6]. In addition, Kazakhstan's recent strategic focus [6] on the integration of renewable energy underscores the importance of accurate wind forecasting to optimize energy production, ensure grid stability, and meet national sustainability goals.

Most recent studies of wind forecasting with deep learning (DL) models were conducted in well-established regions [7, 8, 9] which are already utilizing wind farms, such as China, Europe and North America. The main problem is that Kazakhstan, despite its potential for producing wind energy, has not been the focus of many studies on deep learning wind speed forecasting. The sparsity of available data, climatic conditions, and geographical features becomes a challenge that has not been explored using state-of-the-art models.

It is important to mention that, the Swin Transformer [10] for Time Series (Swin4TS) [3] is a relatively new architecture which was adapted from the computer vision tasks for time series forecasting. Its application for wind forecasting, especially in Kazakhstan, is an unexplored field in the research. The implementation and evaluation of Swin4TS could provide new information on the complex wind patterns and their effective forecasting.

Alongside, my study is focuses on 3 datasets, 2 of them was provided be KazHydroMet [11] and NASAPOWER [12]. The third one was gathered from the Yurt Facility which is located on the territory of Nazarbayev University. All datasets have different sample frequency and the span of the data. Which allows to see broader picture of the underlying wind speed patterns and undercover potential wind speed farm location in the regions which were not considered previously.

Many regional studies rely on the classic statistical methods or basic machine learning techniques. The impact of my research may extend horizons for wind speed forecasting using deep learning on the territory of Kazakhstan for the research community, via collected dataset and wide variety of deep learning models including Transformer and Swin4TS. Additionally, my study focuses not only the wide selection of DL models, but the variety of forecast horizons (FH). It may give a better understanding of the complex wind patterns across the territory of the country. This will allow for expansion of the existing wind farms and construction of the new ones on the promising areas. This will fulfill the nationals goals to reduce the greenhouse gas emissions and provide better overall power grid stability, potentially decreasing the cost for the power in the country.

1.2 Forecast Horizons

In my study I will cover 4 forecast horizons (FH): ultra-short-term, short-term, medium-tern, and long-term.

Forecast Horizon	**Range**	**Applications**
Ultra-short-term	few seconds to 30 minutes ahead	Real-time control of wind turbine in wind farm Real-time trading of energy on energy market Management of grid stability
Short-term	30 minutes to 6 hours ahead	Intra-day trading Balancing of load in power grid system Optimization of the energy storage
Medium-term	6 hours to 1 day ahead	Day-ahead market trading Daily load planning Preventive grid maintenance scheduling
Long-term	more then 1 day	Strategic planning of energy consumption Maintenance scheduling of wind turbines Long-term energy purchasing planning

Table 1.1: Time scale applications in energy systems

1.2.1 Ultra-Short-Term Forecasting

Ultra-Short-Term Forecasting (USTF) (Table 1.1) in general covers very short ranges of time: couple of seconds or minutes, maximum 30 minutes, in some cases it can be extended up to an hour. This instant forecasts primary used for real-time grid operations [13], precise wind turbine control [14], and mechanisms of market clearing [15].

1.2.2 Short-Term Forecasting

Short-Term Forecasting (STF) (Table 1.1) generally ranges from 30 minutes till 6 hours, in rare cases up to a several days. This type of FH plays an important role for load dispatch planning [16], decisions making in load management [17], day-ahead electricity transactions in the market, and ensuring smooth operation of the power grid [18].

1.2.3 Medium-Term Forecasting

Medium-Term Forecasting (MTF) (Table 1.1) ranges from 6 hours till 1 day. This FH crucial for maintenance operations within the power grid, energy trading, decision making in online and offline generations [19].

1.2.4 Long-Term Forecasting

Long-Term Forecasting (LTF) (Table 1.1) is generally forecasts which are more then 1 day. This extended FH aim is strategic scheduling of major maintenance, optimization of long-term operations, general management of wind farms, and planning and construction of the power grids [19].

1.3 Objectives

1. Evaluate the performance of five different deep learning models for wind speed forecasting across different regions of Kazakhstan: 1D Convolutional Neural Network (1D CNN), Long-Short Term Memory (LSTM), WaveNet, Transformer, and Swin Transformer for Time Series (Swin4TS).

2. Compare model performance across four forecast horizons: ultra-short-term (seconds to 30 minutes), short-term (30 minutes to 6 hours), medium-term (6 hours to 1 day), and long-term (more than 1 day).

3. Analyze three different datasets with varying temporal resolutions: NASA

dataset (hourly data), KazHydroMet dataset (3-hour intervals), and Yurt facility dataset (1-minute intervals) to understand wind patterns at different scales.

4. Identify regions in Kazakhstan with more predictable wind patterns that may be suitable for future wind farm development.

5. Assess the potential of the novel Swin4TS model for time series forecasting in the wind energy domain.

6. Provide insights to support Kazakhstan's renewable energy integration and contribute to the research community's understanding of wind speed forecasting in regions with complex geographic features and diverse climatic conditions.

Chapter 2

Literature Review

2.1 Transformer base methods

2.1.1 Attention is all you need

Original Transformer paper [5] represented the first transformer architecture, which relies fully onto attention mechanism, eliminating the need for the recurrent or convolutional neural networks. The self-attention mechanism allows to capture relationships between distant time steps, which can be important in wind speed forecasting. Multi-head attention mechanism extends the capability of self-attention by allowing the model to focus on the different part of the sequence in the same time. Multiple attention heads work in parallel, which allows to capture broader range of the relationships between sequence elements.

2.1.2 Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting

The paper [20] presence an Informer model which addresses long-sequence time-series forecasting via ProbSparse Self-Attention Mechanism, Self-Attention Distilling, and generative style encoding. The use of this model in wind speed forecasting is very promising because of the model's handling long-term dependencies, computational efficiency, and possibility in real-time forecasting. The use of this model offers a promising approach to wind forecasting in Kazakhstan.

2.1.3 Short-term forecasting for multiple wind farms based on transformer model

The paper [21] introduces an application of the Transformer model for the short-term forecasting for multiple wind farms. The encoder-decoder framework was designed without the embedding layers, because time-series forecasting works directly with numerical data rather then categorical tokens which requires embedding. The input itself mapped via linear layer into n-dimensions. The proposed models utilizes the scaled dot-product attention mechanism to capture the relationships in the sequential data, to focus on the key points in time. The dataset was collected from the Northwest Chine throughout the 2018 year. It was split to train (9 month), validation (2 months) and test (1 month) sets.

2.1.4 Long-term Time Series Forecasting with Vision Transformer

The paper [3] shows how the Vision Transformer (ViT) architecture can be applied for the Long-term Time Series Forecasting (LSTF). The proposed model of the Swin transformer was named by authors of the paper as a Swin4TS, which utilizes the window-based attention and hierarchical representation from the original model, but for time-series data. Window-based attention splits the input sequence into windows of fixed-sizes and computes attention within the window. This change reduces the complexity from $O(L^2)$ to $O(L)$. Shifted Window Attention (SWA) allows flow of the information across multiple windows, which is similar to the original's model SWA. Hierarchical representation captures the temporal features at different scales, similar to multi-level feature extraction in computer vision, allowing to understand not only local patterns, but global ones as well.

2.1.5 Multistep short-term wind speed forecasting using transformer

In this paper [22] introduces the sequence-to-sequence transformer models with encoder-decoder architecture, which was adapted for time seies forecasting. The novelty of their paper is Ensemble Empirical Mode Decomposition (EEMD), which decomposes the one dimensional input sequence into 16 dimensional sequence before feeding it into the transformer models. Their dataset contains 999,364 data points, in the period of time from 2002 till 2020. It was split to the train (2002-2018), validation (2019), and test (2020) sets.

2.1.6 A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy

The authors of this paper [23] presents an approach for wind speed and power generation forecasting utilizing Transformer which is enhanced with the wavelet decomposition. Transformer model uses encoder-only architecture for time-series forecasting. In the encoder they use Multilayer Perceptron (MLP), which uses as input temporal information gathered by the encoder. To decompose wind speed sequence into different components they utilized the wavelet transformation, to capture high- and low-frequency signals. This addition improves model's ability to forecast non-stationary patterns in wind speed sequence. Their data was collected in Bahia, Brazil, at different altitudes (100 m, 120 m, and 150 m), with hourly sampling rate. The total size have 744 data points.

2.2 Deep Learning based methods

2.2.1 WaveNet: A Generative Model for Raw Audio

This article [2] presets a WaveNet deep neural network for generating raw audio waveforms. This model is fully probabilistic and auto-regressive, it means that each audio sample produced is based on all the previous audio sampled produced before it. WaveNet's architecture utilizes dilated causal convolutions, which allows it to model long-term sequences without requiring recurrent connections. For our case we adapted WaveNet's architecture for time-series forecasting.

2.2.2 An improved Wavenet network for multi-step-ahead wind energy forecasting

The article [24] proposes a hybrid model called ED-Wavenet-TF, which was designed to enhance multi-step-ahead forecasting in wind energy application. Their model is a encoder-decoder framework which combining two WaveNet networks via multi-head attention mechanism. The datasets which was used in the paper about wind speed and wind power: offshore Australian and inland German wind farms. Each dataset is in total 8760 data points with hourly frequency. The hourly dataset was blueprint and example for gathering related data for the Kazakhstan wind speed forecasting.

2.2.3 One Dimensional Convolutional Neural Network Architectures for Wind Prediction

The article [25] proposes the 1D CNN models which is applicable for real-world scenarios. 1D single CNN (1DS) takes a series of wind speed and wind direction values and forecast separately dominant speed and direction of the wind. 1D Multiple CNN (1DM) is an extension of the 1DS via integration of multiple versions of input, combining multiple instances of 1DS. They utilize several datasets from different places: Stuttgart, Germany and Netherlands. Each data point sampled with 30 minute interval and 30 and 37 years in length respectively.

2.2.4 Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network

In this paper [26] authors present a hybrid deep learning model, which consists of Empirical Wavelet Transform (EWT), Long Short-Term Memory (LSTM), and Elman Neural Network. The proposed novel architecture shows better balance in the empirical adaptability and wavelet-based computation via EWT, and Elman (short term dependencies) and LSTM (long term dependencies) networks makes the model specialized in different frequency components. Their dataset consists of 700 data points (600 : train, 100 : test). In general, proposed model is robust and provides accurate solution for the task of the wind speed forecasting.

2.2.5 Novel wind speed forecasting model based on a deep learning combined strategy in urban energy systems

The study [?] addresses the challenge of the precise short-term wind speed forecasting, particularly in the urban environment. Author propose a hybrid-model which combines Empirical Wavelet Transform (EWT), Sample Entropy (SE), and deep learning (DL) model. EWT is utilized for composition of wind speed data, SE use for selection of relevant components, and the DL for wind speed forecasting. Their method aims to improve the accuracy of wind speed forecasting, in order to achieve better integration of wind energy into the city structure. Due to the non-stationary nature of the wind speed data, EWT and SE performs a important role in feature extraction, and the DL model effectively captures the temporal dependencies.

2.2.6 A Modularity-Enhanced Echo State Network for Nonlinear Wind Energy Predicting

This research [27] proposes a modularized Echo State Network (MESN) to improve the wind speed energy forecasting. The novelty proposed by the authors is in the modular design, divided into: wind speed data decomposition into components of time series; trand patterns clustering (modes-cluster) for the ESN output layer pre-training; wind-turbine clustering (turbines-cluster) to share ESN output matrices. The process starts with decomposition of wind speed dta into trends, seasonal, and residual components. Then, they utilized the K-means clustering for daily patterns of the wind speed and characteristics of the wind turbines. Then, the ESN used for generating preficiton.

2.2.7 Exploring Time Series Models for Wind Speed Forecasting: A Comparative Analysis

The authors of this paper [28] do a comprehensive evaluation of several wind speed forecasting models, spanning statistical, machine learning, and deep learning methods. The aim of this study is to uncover valuable intel for the renewable energy industry, to enhance sustainability and efficiency. Analyzed models: Auto-Regressive Integrated Moving Average, Graphical Model, Linear Regression, Random Forest, Support Vector Regression, Artificial Neural Network, Convolutional Neural Network, Long Short-Term Memory. Findings of the study show the superiority of the deep learning models in wind speed forecasting.

2.2.8 Deep Learning-Based Weather Prediction: A Survey

The survey [29] overviews the application of deep learning architectures in wind speed forecasting. Authors explore models like Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and hybrid models. The paper showed that deep learning (DL) models performs better in the capturing complex non-linear patterns in the wind speed data, then the traditional numerical weather prediction (NWP) models. They also integrated the DL models with NWP, creating a hybrid model that has the strengths of both.

2.2.9 Time Series Forecasting With Deep Learning: A Survey

In this survey [30], authors explored deep learning (DL) architecture utilized for wind speed forecasting. They reviewed common encoder-decoder architectures, which was used for single- and multi-step forecasting. They also highlighted the recent improvements in hybrid DL models, which are combinations of statistical approaches with neural network components. The hybrid models demonstrated performance, which is better then pure statistical and DL methods. However, the survey stated that there are limitations of DL models of handling the time series data which have missing data, or spaced out observations.

2.2.10 A Novel Hybrid Method for Multi-Step Short-Term 70 m Wind Speed Prediction Based on Modal Reconstruction and STL-VMD-BiLSTM

This paper [31] present a hybrid architecture (STL-VMD-BiLSTM) for multi-step short-term wind speed forecasting at elevation of 70 meters. It combines time series decomposition and reconstruction in order to improve the accuracy of the forecast. Decomposition is done using LOESS (STL), which is combination of seasonal and trend decompositions of the original wind speed data. Additionally, they do Variational Mode Decomposition (VMD), to decompose the data into intrinsic model components (IMFs). The paper used the wind data from the wind farm located in Gansu Province in China. They compare their model with STL-BILSTM, VMD-BILSTM, BILSTM, SVR, LGBM, and RF. They state that their models out-performs the other architectures, particularly in capturing characteristics of non-linear data patterns.

Chapter 3

Methodology

3.1 Data Collection

The base for this study starts from the wind data collection, for this I chose 17 cities (Figure 3-1) across the Kazakhstan. These cities include Aktobe, Aktau, Almaty, Astana, Atyrau, Blakhash, Zhezqazghan, Karaganda, Kokshetau, Kostanay, Kyzylorda, Pavlodar, Petropavl, Shymkent, Taldykorgan, Taraz, and Oskemen. The wide range of selected cities cover nearly all geographic regions presented in the Kazakhstan. This will ensure that mine dataset reflects the full representation of the different climate zones and environmental conditions that could be found in the Kazakhstan.

3.1.1 NASA Dataset

The first dataset was acquired from the NASA databases [12], it covers a one-year period of time with the sample rate of 1 hour from 1 January, 2023 till 31 December, 2023. The dataset contains 8760 data points per city. This dataset provides a reliable baseline for analyzing wind patterns in the Kazakhstan. The use of this dataset sourced form the NASA database ensures the accuracy and credibility due to the fact that it was derived from the advanced satellite observations and global climate models.

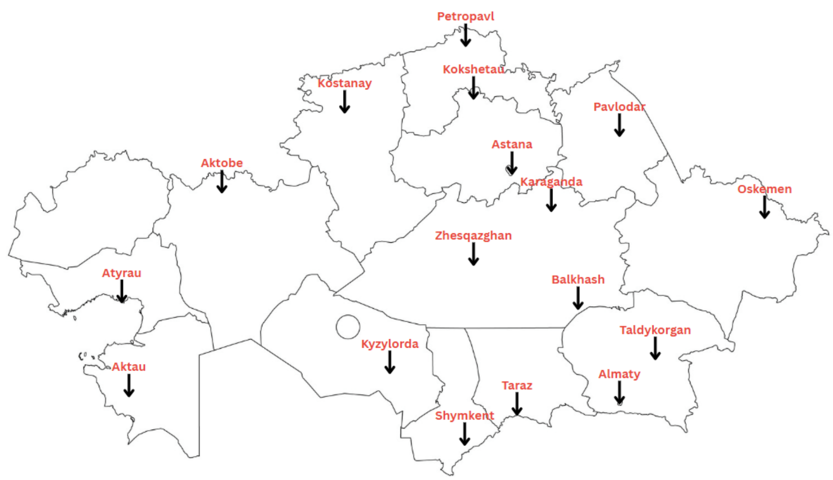

Figure 3-1: Relative map of all listed cities.

3.1.2 KazHydroMet Dataset

The second dataset was acquired from the KazHydroMet [11] database, which serves as crucial repository of meteorological data for Kazakhstan. The range of this dataset is longer from 1 January, 2021 till 31 December, 2024. The sampling rate is 3 hours and each city in the dateset has 11689 data points. This data set will provide a more in-depth study of long-term wind patterns and seasonal variety due to the extended duration. In addition, because this dataset contains more than one year of observations, it helps to lower the effect of anomalies and outliers.

3.1.3 Yurt Dataset

Third dataset was collected from the meteorological station located in Yurt on the territory of Nazarbayev University, with the sample rate of 1 min from 1 January 2022 to 31 December 2023. However, the quality of this data set is lacking due to the unstable sensor work, leading to almost 40% missing data. To address this issue, I implemented a segmentation algorithm. I divided the data set into smaller manageable segments with more than 1000 consecutive data points, which led to approximately 100 usable segments. The Yurt dataset has some challenges in terms of completeness; however, its high resolution might bring valuable insights of very short-term wind patterns.

3.2 Data Description

3.2.1 NASA Dataset Description

NASA dataset (Table A.1) contains 18 cities in Kazakhstan, each location has 8670 data points, which corresponds to hourly data of the 1 year. In the data set, we have columns that represent the year (YEAR), day (DY), hour (HY), month (MO), and wind speed at an altitude of 50 meters (WS50M). The dataset does not contain any missing values for any location. The Box-plot distribution can be observed in Figure B-1.

3.2.2 KazHydroMet Dataset Description

KazHydroMet dataset (Table A.2) contains the same number of cities as NASA dataset, each location has 11688 data points, which corresponds to sample rate of 3 hours of the 3 years. The dataset does not contain any missing values across all locations. In the dataset, we have columns of station name (station), date and time (datetime), and wind speed values (observation). The Box-plot distribution can be observed in Figure B-2.

3.2.3 Yurt Dataset Description

The Yurt dataset (Table A.3) is collected from the yurt facility (Figure 3-2) in Nazarbayev University (Figure 3-2). However, due to the unstable work of meteorological station in the facility the data is not complete and posses almost 40% of missing values (Figure B-7). However, the potential of that facility in the future is very big, because meteorological station located there is capable of collecting weather data with the sampling rate of 1 minute. Despite the fact of missingness, I divided the data into usable chunks, each chunk contains at least 1000 data points. If the full data is used with the missing chunks it will negatively impact the performance of the models, due to the missingness. However, I cannot synthesize the missing chunks based on the previous and upcoming data in the dataset, with the current amount of missing data it will create not trustworthy results. In the Yurt dataset we have columns for: date and time (Time), solar radiation (solar_radiation, Figure B-3, air temperature inside the yurt (temp_inside, Figure B-4), air temperature outside of the yurt (temp_outside, Figure B-5), and wind speed (wind_speed, Figure B-6).

Figure 3-2: Yurt facility on the territory of the Nazarbayev University.

3.3 Setup

For this project I am using PC with this configuration (Table A.4). High-performance HP Z4 G4 Workstation, with the CPU Intel Core i9-7900X and 32 GiB of RAM, ensures effective handling of computations. The GPU RTX 2080 Ti boosts the training process of the deep learning models. The SSD with 1.0 TB storage allows to utilize large datasets with high speeds and efficiency. The Python version 3.12.3, PyTorch version 2.6.0+cu124. Additionally, I utilize libraries: pandas, numpy, matplotlib.pyplot, seaborn.

3.4 Models

My approach is based on utilization of the deep learning model and their forecast of the wind speed based on the historical data for potential places for new wind farms and expansion of the existing ones. For the deep learning models, we are currently using for the other wind forecasting research: 1D CNN, LSTM, WaveNet, Transformer, and Swin4TS. For models evaluation I used metrics:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.1}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.2}$$

3.4.1 1D CNN

The Convolutional Neural Network (CNN) model (Figure 3-3) is specifically designed for wind-speed time-series forecasting. The architecture utilizes the 1D convolutional layers that capture local patterns and dependencies within the temporal data.

Architecture and Key Components:

- **Convolutional Layers:** model utilizes the sequence of convolutional (conv) layers with adjustable kernel sizes to extract temporal features from input
- **Batch Normalization:** Each conv layer is followed by batch normalization to normalize layer inputs, reduce covariate shift.
- **Activation Function:** Rectified Linear Unit (ReLU) activation used after batch normalization to introduce non-linearity.
- **Dropout Regularization:** To prevent overfitting, dropout is applied after in between the layers.
- **Channel Expansion:** To enable the network to learn complex feature representation, the number of channels doubles after each conv layer.

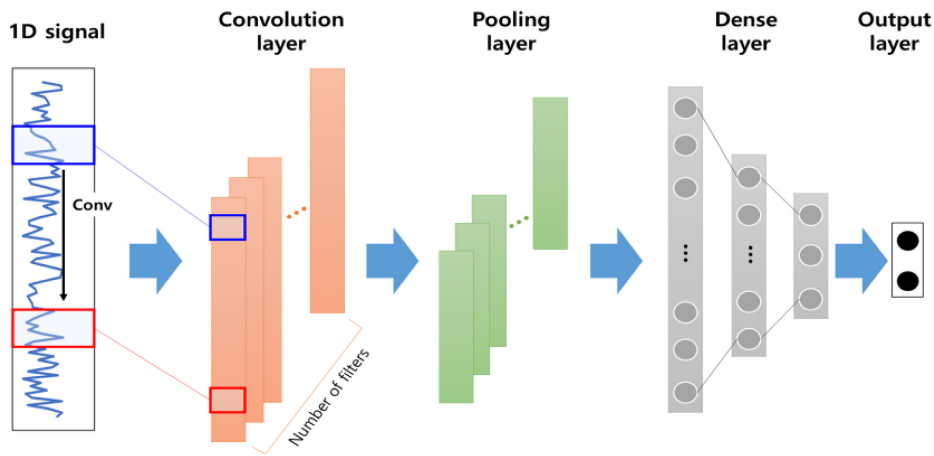

Figure 3-3: Sample CNN architecture for Time-Series Forecasting [1].

- **Global Average Pulling:** To increase robustness to input variations.
- **Fully Connected Layers:** 2 fully connected layers to process pooled features and produce the final output.
- **Weight Initialization:** Xavier initialization used to ensure correct initial scaling and faster cinvergence.

3.4.2 LSTM

The Long Short-Term Memory (LSTM) model (Figure B-8) utilizes the recurrent neural networks (RNN) designed to capture long-range dependencies in time series data.

Architecture and Key Components:

- **LSTM Layers:** Configurable number of stacked LSTM layers to process sequential input data, with tuneable hidden size.
- **Hidden State Initialization:** Ensures consistent behavior, if batch size is changed.
- **Dropout Regularization:** Used between LSTM layers and before the output, to reduce overfitting.

- **State Propagation:** To increase efficiency of modeling of long-term dependencies.
- **Weight Initialization:** Xavier uniform initialization.

3.4.3 WaveNet

This WaveNet model (Figure 3-4) adapts the original architecture that was designed for audio generation for time series forecasting. It utilizes the dilated convolutional layers to model long-term dependencies while preserving the temporal order of the data.

Architecture and Key Components:

- **Dilated Causal Convolutions:** Core component is stack of dilated convolutional layers to capture dependencies at various time scales.
- **Causal Padding:** To ensure that model utilizes only past information for forecasting and maintaines causality.
- **Gated Activation Unit:** Multiplication of sigmopid and tanh activations in order to control flow of the information.
- **Residual Connections:** Each block of the WaveNet model includes residual connection to support gradient flow in training process.
- **Skip Connections:** To allow information to skip some processing steps.
- **Block Structure:** To create a receptive field that grows by exponent with depth.
- **Normalization Layers:** To improve stability and convergence.

3.4.4 Transformer

Adapted version of the original transformer (Figure B-9) attention-based architecture [5] for wind speed time-series forecasting.

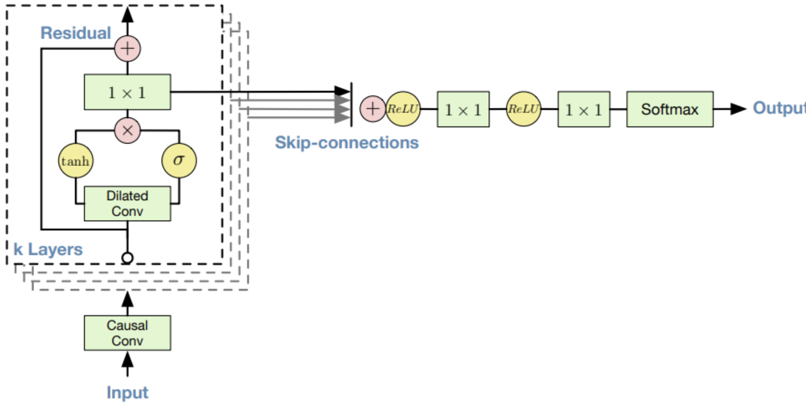

Figure 3-4: WaveNet architecture [2].

Architecture and Key Components:

- **Time Encoding:** Adds sinusoidal positional intel to embedding of the input, to support model understanding of temporal positions.
- **Transformer Encoder:** standard transformer encoder: self-attention mechanism and feed-forward networks.
- **Multi-Head Attention:** Multiple attention heads to pay attention on different parts of the input.
- **Encoder-Only Architecture:** For single-step forecasting.
- **Encode-Decoder Architecture:** For multi-step forecasting.

3.4.5 Swin4TS

The adaptation of the original Swin Vision transformer [10] architecture (Figure 3-5) for time series forecasting [3]. The main features are shifted windows attention mechanism.

Architecture and Key Components:

- **Window-Based Attention:** The model uses attention within a local window range, it was done to reduce computational complexity.

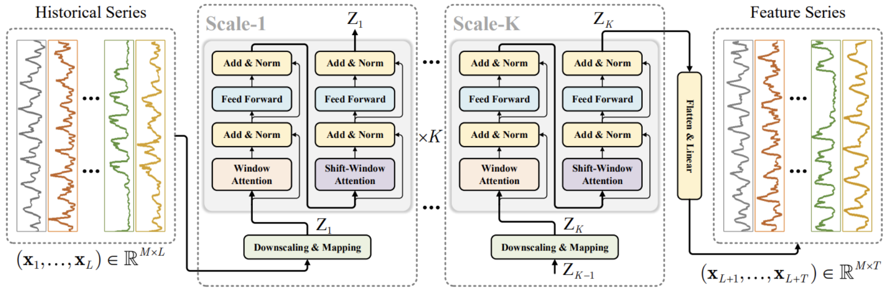

Figure 3-5: Swin4TS architecture [3].

- **Shifted Windows:** Used to connect different local windows with each other.
- **Hierarchical Representation:** To capture the features and different scales.
- **MLP Blocks:** Placed after attention block with GELU activations.

Key Adaptations for Time Series:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.3}$$

$$\text{WindowAttention1D}(x) = \text{Proj}\left(\text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V\right) \tag{3.4}$$

$$x_w = \text{Partition}(x) \in \mathcal{R}^{B\times N_W\times W\times C} \tag{3.5}$$

$$x^{\text{shifted}} = \begin{cases} x, & \text{if shift size} = 0 \\ \text{Roll}(x, -\text{shift size}, \text{dim}=1), & \text{otherwise} \end{cases} \tag{3.6}$$

1. **1D Window Partitioning:** Time series data is segmented into 1D windows:

$$\text{window partition}(x) = x \text{ viewed as } (B, L/W, W, C) \tag{3.7}$$

2. **Shifted Windows:** Cyclic shift along time dimension:

$$\text{shifted x} = \text{roll}(x, \text{-shift size}, \text{dim}=1) \tag{3.8}$$

3. **Attention Stability:** Normalize attention values:

$$\text{attn} = \text{attn} - \max(\text{attn}) \tag{3.9}$$

4. **Length Adaptation:** Padding for arbitrary lengths:

$$\text{pad len} = W - (L \bmod W) \quad \text{if} \quad L \bmod W \neq 0 \quad \text{else} \quad 0 \tag{3.10}$$

3.5 Tuning

To achieve maximum performance and accuracy, I fine-tuned the models using "optuna", the Python package used for hyperparameter tuning. For the fair comparison of performance between models, I fixed the sequence length, according to the forecast horizons, batch size, and number of epochs. The hyperparameter range in which the models tune was selected wide enough to be confident that I found the best solution possible, and short enough to minimize tuning time as much as possible. Pseudocode for hyper parameter tuning:

Algorithm 1 Hyperparameter Tuning for Time Series Forecasting

Require: Model type m, Parameter grid P, Training data D_{train}, Validation data D_{val}, Base config C_{base}

Ensure: Best parameters θ^*, Best model M^*

Initialize $loss^* \leftarrow \infty$
Initialize $\theta^* \leftarrow$ null
Initialize $results \leftarrow []$
for each parameter combination $\theta_i \in P$ **do**
    $C_i \leftarrow C_{base} \cup \theta_i$
    $M_i \leftarrow$ CreateModel(m, C_i)
    Initialize optimizer with learning rate from θ_i
    Initialize scheduler for learning rate adaptation
    $best\_val\_loss \leftarrow \infty$
    $patience\_counter \leftarrow 0$
    $epoch \leftarrow 0$
    while $epoch < max\_epochs$ AND $patience\_counter < patience$ **do**
        $epoch \leftarrow epoch + 1$
        Train model M_i on D_{train} for one epoch
        Compute validation loss val_loss on D_{val}
        if $val\_loss < best\_val\_loss$ **then**
            $best\_val\_loss \leftarrow val\_loss$
            $patience\_counter \leftarrow 0$
            Save model M_i state
        else
            $patience\_counter \leftarrow patience\_counter + 1$
        end if
        Update learning rate using scheduler
    end while
    Record results for parameter combination θ_i
    if $best\_val\_loss < loss^*$ **then**
        $loss^* \leftarrow best\_val\_loss$
        $\theta^* \leftarrow \theta_i$
        $M^* \leftarrow$ Best saved state of M_i
    end if
end for
Save all results and best model to disk
return $\{\theta^*, M^*, loss^*\}$

Chapter 4

Results and Discussion

4.1 Hyper-Parameter Tuning Results

After hyper-parameter tuning on one of the NASA dataset (Aktau), we achieved the results for CNN, LSTM, WaveNet, Transformer, Swin4TS. These hyper-parameters will be used to test performance of the models on other datasets.

4.1.1 1D CNN

The 1D CNN model's hyper-parameter tuning results (Table A.5) shows number of parameters increase with FH from 47681 to 71192. The training time increases with FH from 210 seconds to 836.3 seconds.

4.1.2 LSTM

The LSTM model's hyper-parameter tuning results (Table A.6) shows number of parameters increase with FH from 50397 to 202264. The training time increases with FH respectively from 345.8 seconds to 457.3 seconds.

4.1.3 WaveNet

The WaveNet model's hyper-parameter tuning results (Table A.7) shows number of parameters increase with FH from 31521 to 51832. The training time increases with

FH respectively from 2600.5 seconds to 21464.2 seconds.

4.1.4 Transformer

The Transformers model's hyper-parameter tuning results (Table A.8) displays that number of parameters increase with FH from 25505 to 68632 and training time is increasing with FH from 1751.1 to 2239 seconds.

4.1.5 Swin4TS

The Swin4TS hyper-parameter tuning results (Table A.9) shows that the model maintains the substantial number of parameters !400000 across all FH. The training time is also consistent for all FH, around 450-575 seconds.

4.2 Results on the Datasets

4.2.1 NASA Dataset Results

NASA dataset appears to be the most challenging scenario for wind speed forecasting, all models show high error values. The magnitude of errors increase with respect to FH, that gives a perspective that this dataset contains very complex and harsh wind patterns, which are becoming increasingly difficult to forecast. The root mean square error (RMSE) values increase 3 times from the 1 hour to 24 hour FH for the majority of the models, that means substantial uncertainty at longer FH.

General

I will start with the general overview of the models performance (Figures 4-1, 4-2), which is average RMSE and MAE scores of the model on specific FH across the all 17 cities in NASA dataset (Table A.13).

1D CNN: USTF performance is good, but not exceptional (1 hour RMSE:0.5745). 1DCNN shows relatively to other models steady error increase pattern. On the LTF outperforms Transformer, but still worse than WaveNet. Solid mid-tier in terms of capabilities across all FH and does not show any specific strengths and/or weaknesses at specific FH.

LSTM: Showed exceptional performance in USTF with the lowest RMSE (0.4543) and MAE (0.3091) across all models. However, there is a steep error increase pattern at LTF with RMSE skyrocketing to 3.0718. The relative dominance of the LSTM disappears as the FH increases. Error progression shows linear increase pattern. Despite losing its clear advantage at 12 and 24 hours FH, still keeps competitive performance.

WaveNet: Demonstrated mid-tier performance at USTF (RMSE: 0.5085, MAE: 0.3479). However, showed exceptional performance at longer FH with the best LTF performance (RMSE: 2.8894, MAE: 2.1680). The WaveNet showed the shallowest error increase pattern, which resembles the curve, among all presented models. It is appears that the WaveNet is capable to capture long-term dependencies in the NASA dataset's wind speed patterns then other models. The best performance of the model, is in the MTF.

Transformer: Shows similar performance with WaveNet at 1 hour FH (RMSE: 0.5106). The error curve of the Transformer model is the steepest among all competitive architectures. Reaches the highest RMSE (3.1401) at LTF among all models. The model struggles with maintaining long-term dependencies in this particular dataset. The performance of the model is mid-tier at MTF.

Swin4TS: The clear under-performance of the model at USTF with the highest errors (RMSE:1.0837), however, the performance gap narrows with other models at longer horizons, which indicates insufficient amount of the data for correct forecasting. At LTF, models performance (RMSE: 3.1292) is comparable to other models. The error increase pattern is on the flatter curve than other models. For sufficient performance this model requires longer sequences to utilize the maximum of its architecture.

Details

In this section I will look at some specific details, which might be hidden from the average performance, across the cities from the full NASA dataset and models performance (Table A.10). Observing the data region by region reveals fascinating details that were not observable from the averages. I will try to provide more granular view into critical insights into models' performance across different regions of the Kazakhstan. First of all, I want to define regions with High Forecast Errors and regions with Better Forecast Performance.

High Forecast Errors Regions:

Almaty: consistently shows the highest error scores across all models and FH. For LTF, RMSE ranges from 3.0520 (WaveNet) to 3.4074 (Swin4TS), which is significantly higher then dataset average. This means, Almaty region possesses challenging weather patterns, which is logical, because Almaty is relatively close to the Djungarian gates.

Atyrau: also, shows extremely high error values. For LTF, RMSE values ranges from 3.0080 (WaveNet) to 3.1920 (Transformer), which indicates complex weather dynamics in the western Kazakhstan.

Zhesqasghan: another complex region for forecasting tasks. LTF scores of RMSE ranges from 2.9214 (WaveNet) to 3.3770 (1D CNN). Which indicates some type of unpredictable weather patterns in this region.

Better Forecast Performance Regions:

Taldykorgan: presents consistently lower error values across models, with LFT RMSE ranges from 2.3745 (WaveNet) to 20.5066 (Swin4TS). This is approximately 20% better performance then the other regions.

Taraz: in the same way shows better forecasting capabilities, especially it can be seen for the Transformer model, with LFT RMSE: 2.4821.

Kyzylorda: represents better forecasting capabilities for the USTF and STF, with values lower then 0.50 across all models.

Model-Specific Regional Insights:

1D CNN: Struggles in the Almaty region, however, shows better results in the

Taldykorgan. The performance varies acrossregions and it more pornounce in the MTF then in the other FH.

LSTM: Shows remarkable consistency in USTF, and performance drops at longer FH. According to the data, LSTM performs better in the northernn regions.

WaveNet: Shows the mose consistent performance in LFT across all regions, and shines the most in the areas with continental climate patterns and higher elevation.

Transformer: The most sensitive model among all architectures. Exceptiona performance in the regions: Oskemen and Taraz; but in other regions the performance is poor.

Swin4TS: Performs better at the coastal regions of th Kazakhstan. It underperforms in the USTF, STF, however, it catches up with other models in LFT in some regions.

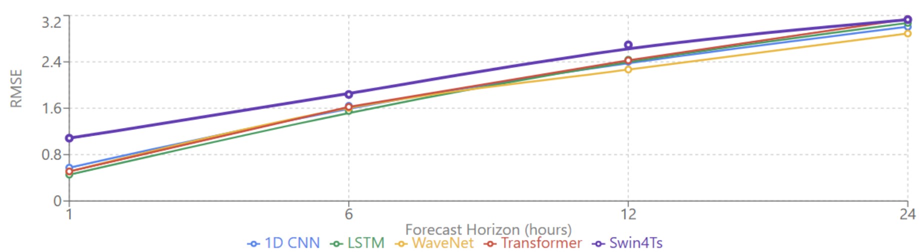

Figure 4-1: RMSE comparison on NASA dataset.

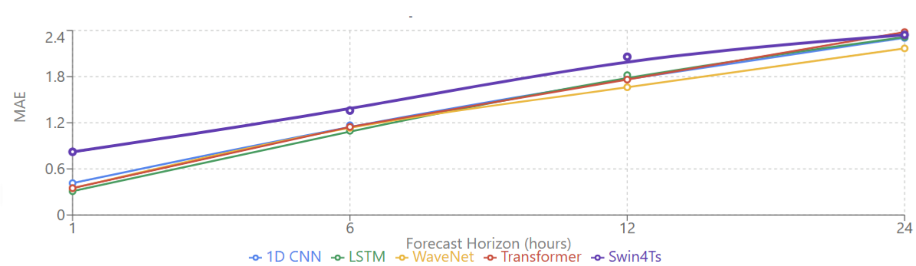

Figure 4-2: MAE comparison on NASA dataset.

4.2.2 KazHydroMet Dataset Results

This dataset shows more moderate error values and a narrower range between the architectures, which indicates that this dataset has mode consistent patterns. The error progression is monotonic curve with less steep then NASA dataset. Potential causes for that - KazHydroMet 3-hour data point frequency collection.

General

I will start with the general overview of the models performance (Figures 4-3, 4-4), which is average RMSE and MAE scores of the model on specific FH across the all 16 cities in KazHydroMet dataset (Table A.14)

1DCNN: Demonstrated good performance, comparable with the LSTM at 1 hour FH (RMSE: 1.4581). The error increase pattern is standard curve. The overall performance is mid-tier across all FH. As in the NASA dataset, 1D CNN did not show any particular strengths or weaknesses at specific horizons. At LTF performed slightly better then LsTM model (RMSE:1.8414).

LSTM: Solid 2-nd place across all forecast horizons, and very competitive at USTF (RMSE: 1.4412, MAE: 1.0674). Maintained a consistent gap in performance with Transformer model. The LTF performance of the LSTM model remains competitive (RMSE: 1.8661).

WaveNet: In comparison with strong performance of this model in NASA dataset, in KazHydroMet dataset WaveNet was under-performing. USTF performance is mid-tier (RMSE: 1.5266). Demonstrated vivid weakness at STF in comparison with other architectures. Placed in the mid-tier in terms of performance across all FH, and at LTF showed average performance (RMSE: 1.8992).

Transformer: Demonstrated exceptional performance across all FH and maintained the lowest RMSE and MAE values. The USTF is very strong (RMSE: 1.2827, MAE: 0.9390). Remained the shallowest error curve, and maintained its superiority even at LTF (RMSE: 1.7265, MAE: 1.2668). Consistent exceptional performance of the model indicates that the KazHydroMet dataset's characteristics are well-suited for the Transformer's architecture.

Swin4TS: Consistent low performance across all FH, and higher error scores in comparison with other models. The model has the steepest error curve. Reaches the highest LTF RMSE (1.9674) among all used architectures. It is clear that dataset's characteristics is not-very suitable for this model.

Details

In this section I will look at some specific details, which might be hidden from the average performance, across the cities from the full KazHydroMet dataset and models performance (Table A.11). In the same manner how NASA dataset uncover some crucial insights fro mthe data, KazHydroMet follows the same trend. The dataset provides significant heterogeneity in forecasting accuracy across all cities, and error patterns are not the same as in the NASA dataset.

High Forecast Errors Regions:

Taraz: Shows as one of the most challenging region to perform wind speed time series forecasting. For LFT RMSE values ranges from 3.0587 (WaveNet) to 3.4399 (1D CNN), this is around 80% higher then in other regions for LFT. This is contrasting with relative better performance of the models in the NASA dataset.

Kokshetau: Also, shows high error values, in particular Swin4TS (LFT RMSE: 2.9995) and 1D CNN (LFT RMSE: 2.3768). The same way as Taraz, this region showed promising results in the NASA dataset.

Atyrau: As well as in the NASA dataset, this region shows worse performance in comparison with others. This can be clearly observed for LFT RMSE values which ranges from 2.3948 (WaveNet) to 2.5746 (Swin4TS). The western region located near Caspian Sea shows high errors across both datasets.

Better Forecast Performance Regions:

Almaty: in the NASA dataset, errors in Almaty regions were higher in comparison with other regions in the same dataset. However, in KazHydroMet dataset this region is among the lowest error metrics across all models and FH. LFT RMSE ranges from 0.8580 (WaveNet) to 0.9564 (Swin4TS).

Shymkent: Also, demonstrates remarkable performance, especially for the Transformer model with LFT RMSE (0.9715). In additional, all models' performance is better in that region, with error values less then half of the dataset's average.

Astana: The capital of the Kazakhstan, surprisingly, shows very good performance with USTF RMSE less then 1.0 across all models.

Model-Specific Regional Insights:

1D CNN: It is weak in Taraz region (USTF RMSE 2.3411) and very competitive in ALmaty region (USTF RMSE: 0.5794). Additionally, shows better performance in the regions located in the southern parts of the Kazakhstan.

LSTM: Outperforms Transformer in Aktau, Atyrau, and Kostanay regions for USTF. Closely competes with Transformer in Petropavl region. Shows better results in the region located in the western parts of the Kazakhstan.

WaveNet: Shows exceptional performance in the south regions of the Kazakhstan. In comparison with Transformer model, weaker in the north regions. Shows overall stable performance across all FH in Shymkent and Aktobe.

Transformer: The best average performance in this dataset, in details, best performance in the most regions. Shows particularly good results in the Alamty (USTF RMSE 0.5413) and Shymkent (USTF RMSE 0.8762). The mose pronounced advantage is in the north regions like Pavlodar (1.2027 RMSE vs next best 1.4114 for WaveNet).

Swin4TS: The best performance shows in the Almaty region (USTF RMSE: 0.5814). However, the gap in performance is more pronounsed in the longer FH in noerth regions of Kazakhstan.

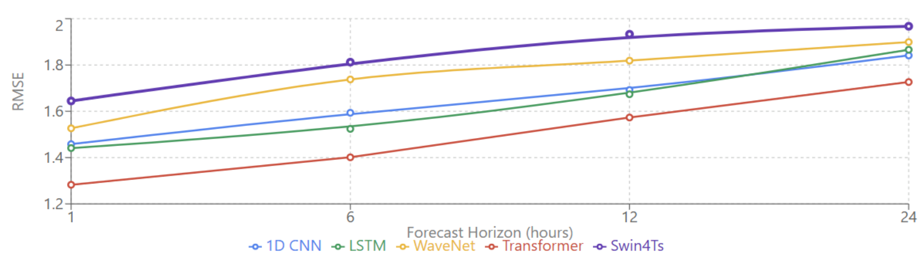

Figure 4-3: RMSE comparison on KazHydroMet dataset.

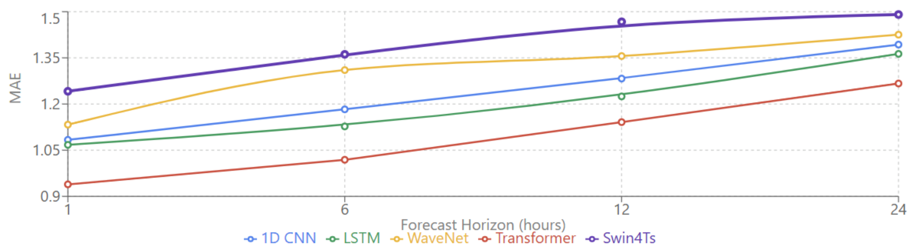

Figure 4-4: MAE comparison on KazHydroMet dataset.

4.2.3 Yurt Dataset Results

The Yurt dataset, which was collected on the territory of Nazarbayev University in the Yurt Facility, showes the most unique error pattern. Some models showed non-monotonic progression. In general, errors are peaking at MTF and then decrease at LTF, which might be an indicator that the data has cyclical patterns. The Yurt dataset reveals significant differences in how models can handle these unique patterns.

General

I will start with the general overview of the models performance (Figures 4-5, 4-6), which is average RMSE and MAE scores of the model on specific FH across the all 17 cities in NASA dataset (Table A.15).

1D CNN: As in the previous datasets, 1D CNN does placed in the mid-tier with average performance. However, for USTF performance is relatively good (RMSE: 0.9638). Similar to other models showed non-monotonic behavior with the peak at MTF (RMSE: 2.0606) and notable drop at LTF (RMSE: 1.8775). This behavior signlas that the model is very sensitive to specific temporal patterns at MTF. Volatile performance across all FH. Captures different aspects of the data patterns at different time-steps.

LSTM: Overall, the best performer across all FH. Excellent USTF performance (RMSE: 0.9108, MAE: 0.6977). Also, shows non-monotonic behavior at MTF with a peak (RMSE: 1.6054) and drop at LTF (RMSE: 1.5068). The same as 1D CNN, LSTM captures cyclical patterns in the data. LSTM demonstrated remarkable adaptability to the dataset's unique patterns.

WaveNet: Showed very competitive to LSTM performance, notably at the MTF. The USTF results of the model is good (RMSE: 0.9527). In comparison with other models, WaveNet shows more monotonic behavior, however, the slope between the MTF and LTF is much gentler. The MTF performance is slightly better (RMSE: 1.5767) then LSTM's performance. Maintained consistent performance across all FH and robustness to the unique patterns in the dataset.

Transformer: Despite clar domination in the KazHydroMet dataset, Transformer demonstrates the bizarre performance across all FH. However, the USTF performance can compete with the results from the LSTM model (RMSE: 0.9491). The deterioration starts at STF (RMSE: 2.1238) and MTF (RMSE: 2.1500), with the partial recovery at LTF (RMSE: 1.7585). Dramatic low performance of the Transformer suggests its sensitivity to some specific temporal patterns in the Yurt dataset.

Swin4TS: Despite low performance in the previous 2 dataset, Swin4TS partially recovers in the Yurt dataset with average performance. The model has less aggressive non-monotonic behavior than 1D CNN. The USTF results (RMSE: 1.0108). Similar to other models, it peaks at MTF (RMSE: 2.0522) and slightly drops at LTF (RMSE: 2.0052). The model shows ability to capture the cyclical patterns, but, in comparison with other models, it is less effective.

Details

In this section I will look at some specific details, which might be hidden from the average performance, across the cities from the full Yurt dataset and models performance (Table A.12). The data from the Yurt dataset, uncovers the temporal patterns of one single location instead of multiple as for last 2 datasets. This gives us more location-specific patterns of forecasting dynamics and model behavior that were not apparent in the previous datasets.

Segment Comparison:

Segment 38191: Shows consistent low errors across all models and FH. For USTF RMSE ranges from 0.8468 (LSTM) to 1.0111 (Swin4TS) and for LFT RMSE ranges from 1.1540 (WaveNet) to 1.6606 (Swin4TS). Remarkable flar progression of the error with FH with only 36% increase from USTF to LTF.

Segment 38181: In comparison with previous segment, this one shows much higher error values. USTF RMSE ranges from 1.1747 (LSTM) to 1.3105 (Swin4TS), and LTF RMSE ranges from 2.1396 (LSTM) to 3.3971 (Swin4TS). Error progression with FH is 82% increase from USTF to LTF.

Model Performance:

1D CNN: decent forecasting capabilities for USTF, and extreme peaks at MTF in both segments. However, model manages to recover in LTF. 1D CNN seems to struggle to forecast in the mid-range FH, and fails to capture the unique temporal patterns in the Yurt location.

LSTM: consistently best performer, it achieves first rank in most segment-horizon combinations. USTF excellence with RMSE: 0.8468, and LFT resilience with RMSE: 1.1396 and 2.1396.

WaveNet: second best and very close to the results of the LTM model, at USTF WaveNet's RMSE: 0.8830 vs LSTM's RMSE: 0.8468. Best LFT performance in the segment 38191 with LFT RMSE: 1.1540.

Transformer: Reasonable forecasting capabilities in the USTF in both segments. Detorioration at STF and MTF, with partial recovery at LTF. However, the difference in performance between segments for this model is more vivid then for any other model.

Swin4TS: The overall performance is similar to transformer and 1D CNN, it manages to outperform Transformer for segment 38191 for USTF, STF, and MTF.

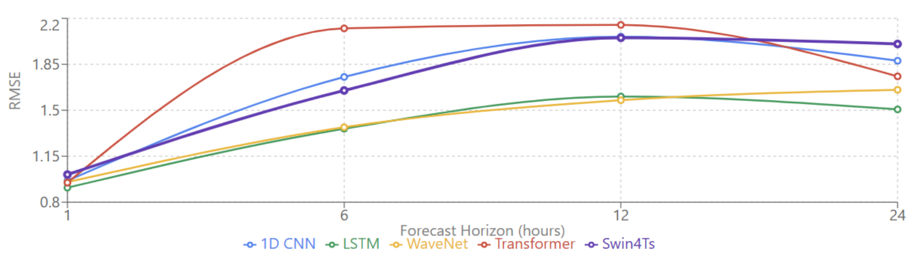

Figure 4-5: RMSE comparison on Yurt dataset.

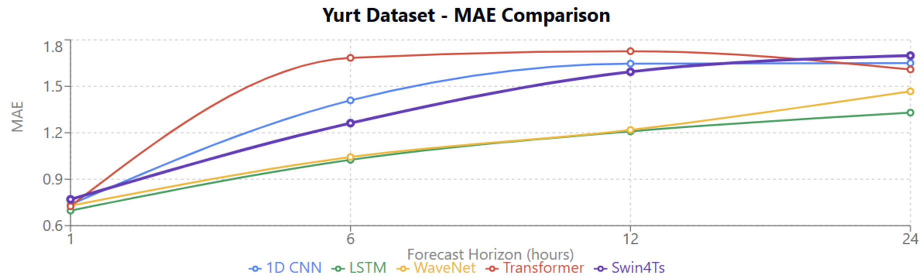

Figure 4-6: MAE comparison on Yurt dataset.

4.3 Discussion

This comprehensive evaluation of 5 time series forecasting models: 1D CNN, LSTM, WaveNet, Transformer, and Swin4TS; across 3 datasets and 4 forecast horizons shows important intel into models' capabilites, limitations, and potential.

Observed pattern across all models and datasets: prediction accuracy deteriorates as the forecast horizon increases from USTF to LTF. This is one of the fundamental challenges of time-series forecasting where uncertainty compounds with temporal distance. Rate of performance degradation is different for each architecture, which gives us important intel into their capabilities.

1D CNN was a solid mid-tier performer across all datasets and FH, without particular strengths and weaknesses, offering a baseline performance with steady error progression.

LSTM showed exceptional forecasting capabilities in USTF on the NASA dataset and the novel Yurt dataset. LSTM achieved lowest error values for USTF. However, for LTF in NASA dataset it's performance fell considerably, which means there are potential limitations in capturing long-range dependencies within this particular dataset. Additionally, it performed competitively in the KazHydroMet, and was out-performed only by the Transformer model.

WaveNet shined in the NASA dataset in terms of capturing long-term dependencies. It showed best LTF performance with a shallowest error curve, which shows the robustness of this architecture to work over longer horizons. In the KazHydroMet dataset, this model was less distinguishable from others. However, in the Yurt dataset, it shined the most. In particular, MTF and LTF showcasing high adaptability.

Transformer is top performer in the KazHydroMet dataset across all FH. This architecture was well suited to the patterns within this dataset. In other hand, Transformer struggled on NASA LTF. On the yurt dataset Transformer was volatile and had a poor performance in STF and MTF. However, the poor performance on the other datasets is might be due to the low resolution of the datasets or the relative small size of the datasets, because the transformer model perform better on the datasets with larger scale.

It is also important to mention the datasets:

NASA dataset appeared as the most challenging one, with high error values and steep error increase with FH across all models, probably due to the reflection of the complex wind patterns across diverse location of Kazakhstan. Showing regions: Taldykorgan, Taraz, and Kyzylorda; for future investigation as potential places for wind power plants as it has the most predictable wind patterns.

KazHydroMet dataset came with moderate error, which suggests more consistent and predictable wind speed patterns, however, it might be influenced by the low resolution of the dataset, data-points were collected every 3 hours. Additionally, this dataset brought some contrast with NASA dataset, especially the Almaty region, which performed poorly in the NASA dataset, here showed good results. Regions like: Almaty, Shymkent, Taraz; were the best for forecasting, which makes them an interesting places for future investigation for potential positions of wind power plant.

Yurt dataset is unique dataset, which was collected on the Yurt Facility on the territory of the Nazarbayev University. This allows us to see location-specific wind patterns with very high data resolution. However, it brought with itself non-monotonic error progression and cyclical patterns. However, LSTM an WaveNet models performed exceptionally well in those conditions.

The Swin Transformer model for Time Series (Swin4TS) model is novel architecture, which was adapted from the computer vision tasks. Across all datasets it perfomed with generally higher metric compred to the established models, especially in USTF. In the NASA dataset it started with the highest errors, however, Swin4TS narrowed the gap considerably in LTF and maintaining relatively flatter error curve, which made him comparable with other models. In the KazHydroMet dataset the picture is relatively similar to the NASA dataset, but in the Yurt dataset Swin4TS showed potential capability for capturing long-term dependencies in the longer FH.

This behavior aligns with the architecture of Vision Transformer (ViT). Generally, ViT requires substantially more data and sequence length to learn and extract insights effectively. The high number of parameters from the hyper-parameter tuning might indicate that this model is more suited for more complex patterns, which might not been present in this study. However, it opens additional direction for further research. I believe the full potential of Swin4TS is yet to be discovered, and in this study I tried to open up the small portion of its capabilities.

4.4 Future Work

In the next studies I want to plan utilize this models for Vietnamese data from the wind power plants, which have much higher resolution, in terms of data-point frequency. Additionally, I want to further explore the capabilities of the Swin4TS model, despite it's poor performance in USTF, STF, and MTF, but it mangaed to close the gap at LTF. Probably this kind of new architecture requires more data and/or multi feature forecasting. I cannot ignore the fact that the parameter size of the Swin4TS was bigger then any other models, but it's training time, also, stayed consistent across FH, which shows it capabilities in terms of training speed. To test it, we can try to rehabilitate the Yurt facility to utilize the powerful stations inside, which additionally opens opportunity for many students to perform their researches with the high quality data that we can collect and process in te]he walls of Nazarbayev university.

4.5 Conclusion

This comprehensive evaluation of five deep learning models—1D CNN, LSTM, WaveNet, Transformer, and Swin4TS—across three datasets and four forecast horizons has yielded valuable insights into wind speed forecasting capabilities in Kazakhstan. The study consistently observed prediction accuracy as forecast horizons increased from USTF to LTF across all models, reflecting the fundamental challenge of compounding uncertainty in time series forecasting. Each model demonstrated distinct strengths and limitations depending on dataset characteristics and forecast horizons. LSTM exhibited exceptional performance in ultra-short-term forecasting, particularly with NASA and Yurt datasets, though its advantage diminished at longer horizons. WaveNet showed remarkable capability in capturing long-term dependencies in the NASA dataset, achieving the best long-term forecast performance with the shallowest error curve among all models. The Transformer model dominated the KazHydroMet dataset across all forecast horizons, suggesting its architecture was particularly well-suited to the patterns within this dataset. Despite occasional underperformance, 1D CNN provided consistent mid-tier results across all scenarios, offering reliable baseline performance. The novel Swin4TS architecture, while generally underperforming at shorter horizons, narrowed the performance gap at longer forecast horizons, suggesting its potential for capturing complex patterns with sufficient data. Its high parameter count indicates it may be better suited for more complex datasets than those used in this study. Regional analysis revealed significant heterogeneity in forecasting accuracy across Kazakhstan. Regions like Taldykorgan, Taraz, and Kyzylorda in the NASA dataset and Almaty, Shymkent, and Astana in the KazHydroMet dataset demonstrated more predictable wind patterns, making them promising candidates for future wind farm development. Interestingly, some regions like Almaty showed contrasting results between datasets, highlighting the importance of multi-source validation in wind resource assessment. This research contributes to the advancement of wind energy integration in Kazakhstan by providing insight into model selection for different forecast horizons and identifying regions with predictable wind patterns. While there was no universally superior model across all datasets and horizons, the findings offer valuable guidance for implementing forecasting systems that can enhance grid stability, optimize energy production scheduling, and support Kazakhstan's renewable energy goals. Future work could explore the capabilities of these models with higher resolution data and investigate multi-feature forecasting approaches, particularly for the promising Swin4TS architecture.

Appendix A

Tables

Table A.1: Summary Statistics for Meteorological Data Across Cities in Kazakhstan from NASA Dataset

Statistic	Aktau	Aktobe	Almaty	Astana	Atyrau	Balkhash	Jazkazgan	Karaganda	Kokshetau	Kostanay	Kyzylorda	Pavlodar	Petropavlovsk	Shymkent	Taldykorgan	Taraz	Ust'-Kamenogorsk
Generic Statistics																	
Count	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000
Missing (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WS50M Statistics																	
Count	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000	8760.000
Max (m/s)	19.020	18.070	15.870	22.650	19.310	17.150	19.480	24.840	23.600	19.000	20.540	23.820	22.860	15.750	15.220	20.610	17.670
Mean (m/s)	7.050	7.061	5.046	6.721	7.256	6.164	6.674	6.743	7.409	6.681	7.043	6.584	7.110	4.912	4.734	4.639	5.341
Min (m/s)	0.080	0.160	0.030	0.210	0.060	0.150	0.080	0.020	0.180	0.060	0.140	0.200	0.140	0.020	0.060	0.050	0.030
Missing (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Std (m/s)	3.357	3.103	2.722	3.231	3.169	2.922	2.857	3.245	3.307	2.916	2.956	2.999	2.952	2.549	2.478	2.722	2.985

Table A.2: Summary Statistics for Meteorological Data Across Cities in Kazakhstan from KazHydroMet Dataset

Statistic	Aktau	Aktobe	Almaty	Astana	Atyrau	Balkhash	Jazkazgan	Karaganda	Kokshetau	Kostanay	Kyzylorda	Pavlodar	Petropavlosk	Shymkent	Taldykorgan	Taraz	Ust'-Kamenogorsk
Generic Statistics																	
Count	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000
Missing (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observation Statistics																	
Count	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 688.000	11 679.000	11 688.000	11 688.000
Max	16.000	22.000	6.000	8.000	21.000	16.000	17.000	12.000	20.000	16.000	20.000	20.000	18.000	18.000	15.000	34.000	16.000
Mean	3.829	2.070	0.575	1.649	4.485	3.818	3.417	2.767	4.006	2.562	2.404	2.558	3.775	1.741	1.409	2.679	2.676
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Missing (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.080	0.000	0.000
Std	2.196	2.220	0.615	1.179	2.861	2.119	2.480	1.773	2.830	1.921	2.422	1.895	2.312	1.568	1.699	3.351	2.517

Table A.3: Summary Statistics for Environmental Data from Yurt Dataset

Statistic	Min	Max	Mean	Median	Std Dev
Generic Statistics					
Count	0.000	15 019.000	2501.696	1701.000	2402.119
Min	0.000	963.000	28.049	0.000	162.730
Max	83.000	1331.000	721.505	749.000	292.052
Mean	6.915	963.000	187.368	138.754	177.134
Std	0.000	392.588	195.503	198.798	103.462
Missing (%)	0.000	100.000	10.435	0.000	30.705
Temperature Inside Yurt					
Count	0.000	15 019.000	2501.696	1701.000	2402.119
Min	13.000	32.000	23.058	24.000	4.719
Max	19.000	37.000	28.087	28.000	3.871
Mean	16.568	33.752	25.623	26.656	3.949
Std	0.000	5.738	1.421	1.178	0.953
Missing (%)	0.000	100.000	10.435	0.000	30.705
Temperature Outside					
Count	0.000	15 019.000	2501.696	1701.000	2402.119
Min	−68.000	26.000	0.035	2.600	14.443
Max	−15.000	37.000	12.687	16.000	13.568
Mean	−18.922	29.860	7.010	9.478	12.907
Std	0.000	9.148	3.267	3.318	1.505
Missing (%)	0.000	100.000	10.435	0.000	30.705
Wind Speed					
Count	0.000	15 019.000	2501.696	1701.000	2402.119
Min	0.000	7.600	0.252	0.000	1.283
Max	3.100	22.000	8.733	8.000	3.411
Mean	0.500	7.600	2.482	2.277	1.363
Std	0.000	3.132	1.414	1.340	0.581
Missing (%)	0.000	100.000	10.435	0.000	30.705

Hardware Information	
Model	HP HP Z4 G4 Workstation
Memory	32.0 GiB
Processor	Intel® Core™ i9-7900X $\times$ 20
Graphics	NVIDIA GeForce RTX™ 2080 Ti, 11GiB VRAM
Disk Capacity	1.0 TB

Table A.4: PC Hardware Specifications

Table A.5: CNN Hyper Parameter Tuning Results by Forecast Horizon

Model	Forecast Horizon	Sequence Length	RMSE	MAE	MAPE	Parameters	Training Time (s)	Best Hyperparameters
cnn	1	12	0.4461	0.2762	6.27	47,681	210.0	hidden_dim: 128 kernel_size: 3 learning_rate: 0.001 num_layers: 3
cnn	6	24	1.1657	0.8132	17.61	68,870	269.5	hidden_dim: 128 kernel_size: 5 learning_rate: 0.001 num_layers: 3
cnn	12	48	1.3403	0.9641	22.27	69,644	358.8	hidden_dim: 128 kernel_size: 5 learning_rate: 0.001 num_layers: 3
cnn	24	96	1.4421	1.0696	24.40	71,192	836.3	hidden_dim: 128 kernel_size: 5 learning_rate: 0.001 num_layers: 3

Table A.6: LSTM Hyper Parameter Tuning Results by Forecast Horizon

Model	**Forecast Horizon**	**Sequence Length**	**RMSE**	**MAE**	**MAPE**	**Parameters**	**Training Time (s)**	**Best Hyperparameters**
lstm	1	12	0.4519	0.2808	6.65	50,497	345.8	dropout: 0.2 hidden_size: 64 learning_rate: 0.001 num_layers: 2
lstm	6	24	0.8343	0.5790	12.43	199,942	420.1	dropout: 0.2 hidden_size: 128 learning_rate: 0.001 num_layers: 2
lstm	12	48	0.6452	0.4557	10.47	200,716	440.8	dropout: 0.0 hidden_size: 128 learning_rate: 0.001 num_layers: 2
lstm	24	96	0.6624	0.4785	10.63	202,264	457.3	dropout: 0.0 hidden_size: 128 learning_rate: 0.001 num_layers: 2

Table A.7: WaveNet Hyper Parameter Tuning Results by Forecast Horizon

Model	Forecast Horizon	Sequence Length	RMSE	MAE	MAPE	Parameters	Training Time (s)	Best Hyperparameters
wavenet	1	12	0.4442	0.2807	6.53	31,521	26040.5	kernel_size: 3 learning_rate: 0.001 num_blocks: 2 num_layers: 2 residual_channels: 32 skip_channels: 16
wavenet	6	24	1.6131	1.0729	23.93	51,238	19826.3	kernel_size: 3 learning_rate: 0.0001 num_blocks: 2 num_layers: 3 residual_channels: 32 skip_channels: 32
wavenet	12	48	2.2496	1.6335	39.52	47,292	21464.2	kernel_size: 3 learning_rate: 0.001 num_blocks: 2 num_layers: 3 residual_channels: 32 skip_channels: 16
wavenet	24	96	2.6828	1.9914	46.13	51,832	19859.4	kernel_size: 3 learning_rate: 0.001 num_blocks: 2 num_layers: 3 residual_channels: 32 skip_channels: 32

Table A.8: Transformer Hyper Parameter Tuning Results by Forecast Horizon

Model	Forecast Horizon	Sequence Length	RMSE	MAE	MAPE	Parameters	Training Time (s)	Best Hyperparameters
transformer	1	12	0.4932	0.2996	6.74	25,505	1751.1	d_model: 32 dim_feedforward: 128 dropout: 0.1 learning_rate: 0.001 num_decoder_layers: 0 num_encoder_layers: 2 num_heads: 4
transformer	6	24	1.6060	1.0813	23.68	67,462	1893.6	d_model: 64 dim_feedforward: 128 dropout: 0.1 learning_rate: 0.001 num_decoder_layers: 2 num_encoder_layers: 2 num_heads: 2
transformer	12	48	1.7740	1.2787	29.22	51,340	2205.4	d_model: 64 dim_feedforward: 64 dropout: 0.1 learning_rate: 0.001 num_decoder_layers: 2 num_encoder_layers: 2 num_heads: 4
transformer	24	96	1.4934	1.1115	27.15	68,632	2239.0	d_model: 64 dim_feedforward: 128 dropout: 0.1 learning_rate: 0.001 num_decoder_layers: 2 num_encoder_layers: 2 num_heads: 4

Table A.9: Swin4TS Hyper Parameter Tuning Results by Forecast Horizon

Model	Forecast Horizon	Sequence Length	RMSE	MAE	MAPE	Parameters	Training Time (s)	Best Hyperparameters
swin	1	12	1.2226	0.9134	21.06	400,385	515.3	attn_drop_rate: 0.1 depths: [2, 2, 2, 2] drop_rate: 0.1 embed_dim: 64 learning_rate: 0.001 num_heads: [2, 4, 8, 16] window_size: 4
swin	6	24	2.1879	1.6598	37.55	400,966	515.5	attn_drop_rate: 0.1 depths: [2, 2, 2, 2] drop_rate: 0.1 embed_dim: 64 learning_rate: 0.001 num_heads: [2, 4, 8, 16] window_size: 4
swin	12	48	2.6313	2.0530	51.81	401,868	450.8	attn_drop_rate: 0.1 depths: [2, 2, 2, 2] drop_rate: 0.1 embed_dim: 64 learning_rate: 0.001 num_heads: [2, 4, 8, 16] window_size: 4
swin	24	96	1.7987	1.3590	32.23	402,648	573.5	attn_drop_rate: 0.1 depths: [2, 2, 2, 2] drop_rate: 0.1 embed_dim: 64 learning_rate: 0.001 num_heads: [2, 4, 8, 16] window_size: 4

Table A.10: Comparison of Different Models on NASA Dataset

NASA Dataset	Metric	1D CNN				LSTM				WaveNet				Transformer				Swim4TS			
		1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours
Aktau	RMSE	0.8772	1.9145	3.0981	3.6502	0.4520	1.8817	3.0177	3.9405	0.5207	1.8926	2.7086	3.6609	0.5253	1.9135	2.8296	3.9036	1.2221	2.1579	3.4865	3.8838
	MAE	0.4949	1.3595	2.3339	2.6408	0.2808	1.2758	2.0177	2.9502	0.3547	1.3799	1.9976	2.7731	0.3394	1.3355	2.0609	3.0388	0.6634	1.5630	2.6952	3.1690
Aktobe	RMSE	0.4539	1.4549	2.2354	3.0534	0.4129	1.4019	2.7144	3.2428	0.4463	1.4425	2.2062	2.7806	0.4587	1.4876	2.2934	3.0305	0.8935	1.7623	2.9984	3.1994
	MAE	0.3377	1.0393	1.6570	2.2632	0.2955	0.9847	1.5984	2.4887	0.3272	1.0151	1.6177	2.0774	0.3303	1.0629	1.6794	2.2916	0.5862	1.3000	2.0205	2.5555
Almaty	RMSE	0.7254	2.2599	2.9774	3.1158	0.6314	2.1239	3.3203	3.2550	0.7015	2.1702	2.7064	3.0520	0.6823	2.1122	2.6218	3.3116	1.5672	2.3838	3.0492	3.1915
	MAE	0.5565	1.6519	2.2762	2.4465	0.4655	1.5711	2.4624	2.5127	0.5268	1.5870	2.0708	2.3584	0.5003	1.5522	2.1906	2.5262	1.2423	1.8358	2.4276	2.5914
Astana	RMSE	0.6090	1.7249	2.4860	3.6656	0.4561	1.5501	2.4213	3.4250	0.5185	1.4469	2.4154	3.1742	0.5080	1.6342	2.5213	3.2974	1.1042	1.8516	2.8260	3.4304
	MAE	0.4263	1.1992	1.8195	2.7296	0.3054	1.0582	1.7232	2.5337	0.3766	1.1487	1.7765	2.3632	0.3400	1.1240	1.7967	2.4407	0.7971	1.3406	2.0846	2.6045
Atyrau	RMSE	0.5262	1.5332	2.4077	3.1282	0.3996	1.4746	2.2252	3.3499	0.4439	1.4796	2.3057	3.0816	0.4692	1.4916	2.2950	3.2723	0.9601	1.7856	2.8503	3.4101
	MAE	0.3862	1.1029	1.8268	2.4243	0.2843	1.0475	1.6231	2.7173	0.3106	1.0687	1.7149	2.3570	0.3606	1.0860	1.7129	2.5600	0.7358	1.3431	2.2128	2.7679
Balkhash	RMSE	0.4765	1.6480	2.5536	2.7593	0.4666	1.5483	2.6319	3.0080	0.4719	1.5589	2.2576	2.7330	0.4642	1.5266	2.3615	3.1192	1.0655	1.7321	2.7748	2.7718
	MAE	0.3450	1.2058	1.9394	2.1273	0.2970	1.0934	1.8727	2.2900	0.3319	1.1166	1.6789	2.0880	0.3438	1.1000	1.7559	2.4510	0.8063	1.3110	2.1067	2.2257
Zhezkazghan	RMSE	0.5833	1.5472	2.2828	3.0370	0.4360	1.4998	2.4736	2.9003	0.4761	1.5021	2.1887	2.8214	0.4557	1.5342	2.5678	2.7334	1.0684	1.7260	2.6390	2.7760
	MAE	0.4484	1.0668	1.6506	2.3587	0.3067	1.0404	1.7856	2.1712	0.3351	1.0526	1.5890	2.1165	0.3211	1.0654	1.8533	2.0792	0.7596	1.2450	1.9550	2.4899
Karaganda	RMSE	0.7425	1.7013	2.5990	4.1612	0.4618	1.5255	2.6668	3.4428	0.5751	1.6731	2.5704	3.5663	0.5489	1.7891	2.9071	3.6993	1.0928	1.9371	3.0784	4.0386
	MAE	0.5317	1.2048	1.9100	3.1425	0.3159	1.0581	1.9080	2.4798	0.3974	1.1735	1.8241	2.6299	0.3558	1.2204	2.0330	2.7201	0.8106	1.3934	2.3173	3.0523
Kokshetau	RMSE	0.6276	1.6230	2.3838	3.0315	0.4504	1.5079	2.1163	3.0721	0.6112	1.5537	2.2648	3.0059	0.5972	1.5929	2.4129	3.2022	1.0908	1.9275	2.7289	3.2363
	MAE	0.4581	1.1280	1.7325	2.2659	0.3006	1.0270	1.5601	2.2894	0.3809	1.0652	1.6437	2.2297	0.3513	1.0946	1.7398	2.3951	0.8130	1.3776	2.0825	2.3960
Kostanay	RMSE	0.5299	1.5589	2.2779	2.6151	0.4364	1.5089	2.1246	2.7122	0.5072	1.5223	2.1403	2.7963	0.4936	1.5590	2.2788	3.2001	1.0904	1.7832	2.4987	2.9860
	MAE	0.3554	1.0432	1.6223	1.9660	0.2916	1.0227	1.4886	2.1014	0.3365	1.0484	1.5321	2.0515	0.3359	1.0731	1.6328	2.3448	0.8193	1.2958	1.8598	2.2944
Kyzylorda	RMSE	0.5127	1.5067	2.2826	2.9201	0.4391	1.4824	2.2679	2.8397	0.4690	1.5321	2.1126	2.5889	0.4788	1.5467	2.2094	2.8810	1.0326	1.7338	2.5990	2.7805
	MAE	0.3792	1.0947	1.7255	2.2995	0.3058	1.0617	1.6633	2.1747	0.3236	1.1000	1.5697	1.9984	0.3379	1.1228	1.6520	2.2430	0.7978	1.3124	2.0436	2.2139
Shymkent	RMSE	0.5059	1.6157	2.0620	2.4479	0.4676	1.6040	2.2106	2.4541	0.4630	1.6203	2.1088	2.4398	0.4907	1.5755	2.3122	2.5926	1.1352	1.7856	2.2208	2.4992
	MAE	0.3576	1.1656	1.5727	1.9332	0.3193	1.1437	1.6191	1.8982	0.3274	1.1483	1.5599	1.8697	0.3421	1.2170	1.7251	1.9631	0.8707	1.3397	1.7381	1.9669
Taldykorgan	RMSE	0.5371	1.5198	2.0482	2.3823	0.4836	1.5274	2.3691	2.6078	0.5179	1.5191	1.9961	2.3745	0.5184	1.5319	2.1549	2.6331	1.0273	1.7492	2.2430	2.5066
	MAE	0.3991	1.1135	1.5581	1.8852	0.3044	1.1020	1.8173	2.0586	0.3340	1.0936	1.4967	1.8419	0.3385	1.0955	1.6150	2.0339	0.7944	1.3216	1.7854	2.0608
Taraz	RMSE	0.5012	1.7162	2.2180	2.4635	0.4990	1.6674	2.4493	2.8321	0.5011	1.7355	2.2287	2.5190	0.5185	1.7854	2.3579	2.9843	1.1878	1.8507	2.4931	2.7288
	MAE	0.3542	1.1902	1.5937	1.8476	0.3234	1.1984	1.7156	1.9484	0.3447	1.2038	1.5976	1.8637	0.3621	1.2274	1.6236	2.2117	0.8345	1.3611	1.8284	2.0573
Oskemen	RMSE	0.5590	1.4322	2.2617	2.8385	0.4181	1.3869	2.5792	3.4054	0.4346	1.3971	2.0851	2.8812	0.4638	1.3712	2.2745	3.3218	0.9495	1.6240	2.6640	3.1349
	MAE	0.4085	1.0283	1.7133	2.2123	0.2889	0.9872	1.8714	2.6059	0.3052	1.0022	1.5553	2.1610	0.3351	0.9774	1.6792	2.5655	0.7404	1.2185	2.0118	2.5365
Pavlodar	RMSE	0.6706	1.7441	2.4108	3.0700	0.4469	1.4952	2.7720	3.0347	0.4754	1.5481	2.1900	2.9226	0.4848	1.5641	2.4412	3.1420	1.0060	1.7885	2.6856	3.1567
	MAE	0.4754	1.2216	1.6874	2.2702	0.2952	1.0064	1.6230	2.2704	0.3158	1.0759	1.5822	2.1587	0.3247	1.0684	1.7179	2.3401	0.7555	1.2962	1.9490	2.4279
Petropavl	RMSE	0.5032	1.4652	2.2573	2.7566	0.4035	1.4176	2.0872	2.8983	0.4952	1.4945	2.0153	2.6050	0.4803	1.4712	2.3011	3.0974	0.9394	1.6666	2.3757	3.0959
	MAE	0.3590	1.0395	1.6587	2.0813	0.2752	0.9793	1.4842	1.9856	0.3211	1.0400	1.4676	1.9230	0.3199	1.0399	1.6659	2.2423	0.7550	1.2519	1.8170	2.3564
Average	RMSE	0.5745	1.6421	2.4031	3.0057	0.4443	1.5626	2.4461	3.0718	0.4885	1.6052	2.2461	2.8844	0.5106	1.6235	2.4259	3.1401	1.0837	1.8340	2.6839	3.1292
	MAE	0.4149	1.1679	1.7534	2.3050	0.3051	1.0957	1.8155	2.3156	0.3478	1.1342	1.6627	2.1680	0.3492	1.1463	1.7620	2.3798	0.8222	1.3594	2.0586	2.4540

Table A.11: Comparison of Different Models on KazHydroMet Dataset

KazHydroMet Dataset	Metric	1D CNN				LSTM				WaveNet				Transformer				Swim4Ts			
		1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours
Aktau	RMSE	1.5477	1.7040	1.8100	1.8518	1.5351	1.6538	1.8043	2.0382	1.6169	1.8255	1.9846	1.9793	1.5791	1.4915	1.6297	1.7488	1.7232	1.9308	2.0129	2.0159
	MAE	1.1641	1.2882	1.3634	1.4184	1.1540	1.2474	1.3460	1.5062	1.2215	1.3924	1.4257	1.5219	1.0274	1.1167	1.2317	1.3249	1.3326	1.4909	1.5581	1.5575
Aktobe	RMSE	1.6732	1.7601	1.9387	2.0832	1.6401	1.7182	1.8089	2.1176	1.6755	1.9484	1.9928	2.1231	1.5211	1.6287	1.8615	2.0285	1.8739	2.1070	2.2116	2.1671
	MAE	1.0373	1.1032	1.2201	1.2985	1.0232	1.0774	1.1220	1.3328	1.0860	1.2241	1.2666	1.3608	0.9002	0.9552	1.1045	1.1994	1.2256	1.3520	1.4599	1.4330
Almaty	RMSE	0.5704	0.5799	0.5793	0.5874	0.5623	0.5861	0.5917	0.6142	0.5865	0.5881	0.5875	0.5860	0.5543	0.5593	0.5721	0.6091	0.5814	0.5893	0.5989	0.5964
	MAE	0.4790	0.4815	0.4900	0.4951	0.4731	0.4774	0.4777	0.4901	0.4779	0.4774	0.4827	0.4837	0.4251	0.4370	0.4556	0.4975	0.5046	0.5127	0.5137	0.5086
Astana	RMSE	0.7584	0.8175	0.8413	0.8874	0.7466	0.7775	0.8036	0.8532	0.7939	0.8558	0.8936	0.9434	0.6825	0.7211	0.8101	0.8826	0.8155	0.8966	0.9330	0.9615
	MAE	0.5774	0.6173	0.6426	0.7092	0.5664	0.5940	0.6238	0.6505	0.6151	0.6571	0.6880	0.7312	0.5225	0.5484	0.6130	0.6640	0.6348	0.6945	0.7430	0.7552
Atyrau	RMSE	1.7394	1.8444	2.1329	2.3744	1.7665	1.8437	2.1419	2.6037	1.8419	2.1954	2.3238	2.3948	1.6424	1.8449	1.9268	2.1110	2.0165	2.3082	2.4501	2.5646
	MAE	1.3388	1.5096	1.6956	1.9379	1.3474	1.4077	1.6199	1.9557	1.4333	1.6980	1.8085	1.9120	1.1718	1.2507	1.4380	1.6011	1.5913	1.8054	1.9586	2.0921
Balkhash	RMSE	1.7812	1.9413	2.0990	2.1016	1.7564	1.8836	2.1578	2.0027	1.9089	2.0493	2.1909	2.1490	1.6055	1.7366	1.9705	2.0838	1.8954	2.0347	2.1511	2.2289
	MAE	1.3483	1.4737	1.5582	1.6399	1.3251	1.4295	1.5762	1.5311	1.4413	1.5635	1.6206	1.6618	1.1956	1.2983	1.4790	1.5881	1.4585	1.5831	1.6796	1.7289
Zhezkazghan	RMSE	1.5404	1.7637	1.9399	2.1307	1.5013	1.6407	1.8925	2.2308	1.5658	1.8495	2.1106	2.2230	1.2939	1.5224	1.8376	2.0762	1.8862	2.0444	2.3616	2.3438
	MAE	1.1251	1.2920	1.4318	1.5915	1.1017	1.1848	1.3618	1.6272	1.1670	1.4246	1.5784	1.6782	0.9254	1.0893	1.3246	1.5003	1.4129	1.5336	1.7829	1.7870
Karaganda	RMSE	1.3209	1.4189	1.0976	1.6238	1.3171	1.3651	1.5230	1.6669	1.4007	1.5261	1.6000	1.6810	1.1377	1.2280	1.4070	1.5402	1.5100	1.6693	1.7522	1.8081
	MAE	1.0185	1.0976	1.1698	1.2537	1.0107	1.0509	1.1618	1.3239	1.0912	1.1775	1.2385	1.2977	0.8727	0.9410	1.0610	1.1597	1.1814	1.3157	1.3712	1.4307
Kokshetau	RMSE	1.6438	2.0530	2.2293	2.3768	1.8346	1.9477	2.1772	2.3023	1.9669	2.2035	2.4378	2.6139	1.5878	1.7275	2.0234	2.2614	2.0739	2.3175	2.5395	2.5995
	MAE	1.3991	1.5978	1.7135	1.8617	1.3821	1.4694	1.6411	1.7322	1.4489	1.7099	1.8745	2.0241	1.1890	1.3012	1.4884	1.6318	1.5937	1.7876	1.9936	2.1288
Kostanay	RMSE	1.1640	1.2881	1.3529	1.4239	1.1619	1.2048	1.2502	1.4478	1.1951	1.3664	1.4426	1.6085	0.9952	1.1059	1.1978	1.3474	1.3337	1.4927	1.6284	1.6357
	MAE	0.8910	0.9993	1.0469	1.1164	0.8929	0.9230	0.9647	1.1069	0.9315	1.0463	1.1222	1.2604	0.7864	0.8451	0.9172	1.0276	1.0475	1.1737	1.3169	1.3156
Kyzylorda	RMSE	1.5894	1.7257	1.9881	2.1199	1.5568	1.6637	1.7919	2.0364	1.6537	1.9834	2.0672	2.2511	1.4229	1.5747	1.7655	1.8492	1.9147	2.1605	2.2649	2.3777
	MAE	1.1353	1.2335	1.4266	1.5259	1.1280	1.1703	1.2475	1.3504	1.1763	1.3799	1.4547	1.6110	0.9892	1.0731	1.2300	1.3060	1.3838	1.5356	1.6589	1.7138
Shymkent	RMSE	0.9907	1.0503	1.0745	1.0723	0.9616	0.9955	1.0245	1.0919	1.0353	1.0920	1.0980	1.1079	0.8782	0.9076	0.9241	0.9715	0.9973	1.0585	1.0873	1.0859
	MAE	0.7292	0.7618	0.7894	0.7867	0.6999	0.7438	0.7584	0.7925	0.7448	0.7926	0.8099	0.8219	0.6480	0.6722	0.6887	0.7115	0.7419	0.7682	0.8038	0.7858
Taraz	RMSE	2.3341	2.6415	2.9293	3.4399	2.3137	2.5244	2.7205	2.9847	2.3774	2.9058	3.1207	3.0587	2.0620	2.1740	2.4108	2.8003	2.7542	2.9401	3.2023	3.1457
	MAE	1.7415	1.8993	2.1146	2.5656	1.6930	1.8042	1.8693	2.0388	1.7381	2.0217	2.1774	2.1417	1.5271	1.5728	1.7015	1.9310	1.9657	2.0855	2.2240	2.1773
Oskemen	RMSE	1.5585	1.6973	1.7567	1.7879	1.5471	1.6205	1.6859	2.1596	1.7037	1.8324	1.8659	1.9244	1.4403	1.5714	1.6690	1.7636	1.6624	1.8272	1.9226	1.9353
	MAE	1.1525	1.2769	1.3299	1.3472	1.1454	1.2224	1.2946	1.6411	1.2590	1.3324	1.4171	1.4675	1.0500	1.1465	1.2432	1.3186	1.2514	1.3815	1.4644	1.4736
Pavlodar	RMSE	1.3524	1.4420	1.5848	1.6411	1.3188	1.3699	1.5216	1.7820	1.4414	1.5949	1.6216	1.8761	1.2027	1.3256	1.4659	1.6150	1.5310	1.6736	1.7173	1.7519
	MAE	0.9912	1.0349	1.1244	1.2163	0.9577	0.9989	1.1172	1.3026	1.0308	1.1588	1.1842	1.2276	0.8481	0.9195	1.0005	1.1116	1.1238	1.2278	1.2873	1.3083
Petropavl	RMSE	1.5642	1.6719	1.8142	1.9511	1.5360	1.5889	1.7877	1.9277	1.6495	1.9270	1.9153	2.0627	1.3184	1.4811	1.6892	1.9353	1.7817	1.9488	2.0724	2.1187
	MAE	1.1261	1.2888	1.4052	1.5270	1.1866	1.2361	1.3758	1.4580	1.2609	1.4105	1.4866	1.6056	1.0231	1.1358	1.2834	1.4752	1.4119	1.5348	1.6617	1.6570
Average	RMSE	1.4501	1.5937	1.6921	1.8414	1.4412	1.5241	1.6737	1.8661	1.5266	1.7377	1.8189	1.8992	1.2827	1.4013	1.5753	1.7265	1.6446	1.8123	1.9331	1.9674
	MAE	1.0837	1.1828	1.2828	1.3932	1.0674	1.1273	1.2248	1.3631	1.1331	1.3104	1.3500	1.4286	0.9390	1.0255	1.1621	1.2668	1.2415	1.3717	1.4873	1.5089

Table A.12: Comparison of Different Models on Yurt Dataset

Yurt Dataset	Metric	1D CNN				LSTM				WaveNet				Transformer				Swim4Ts			
		1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours	1 hour	6 hours	12 hours	24 hours
Segment 38191	RMSE	0.7057	1.1357	1.2250	1.7272	0.6468	0.8558	1.1187	1.3496	0.6830	0.9177	1.0918	1.1540	0.7036	0.9906	1.4195	1.9315	0.7111	0.8812	0.9949	1.2572
	MAE	0.5251	0.8886	0.9668	1.4318	0.4857	0.6508	0.8211	0.9244	0.5142	0.6943	0.7530	0.8624	0.5553	1.2258	1.0692	0.8390	0.5337	0.5115	0.7845	0.9606
Segment 38181	RMSE	1.2218	2.3721	2.8962	2.4823	1.1747	1.8636	2.0921	2.1396	1.2223	1.8244	2.1415	2.3551	1.1847	2.0569	2.3806	2.7851	1.3105	2.6211	3.1064	3.3671
	MAE	0.9487	1.9305	2.3266	1.8688	0.9088	1.4004	1.5972	1.7360	0.9442	1.3930	1.6838	2.0731	0.8961	2.1418	2.3844	2.3786	1.0098	2.0136	2.4032	2.4358
Average	RMSE	0.9638	1.7539	2.0606	1.8775	0.9108	1.3597	1.6054	1.5068	0.9527	1.3711	1.5767	1.6561	0.9491	2.1238	2.1800	1.7585	1.0108	1.6511	2.0522	2.0052
	MAE	0.7369	1.4095	1.6467	1.6503	0.6977	1.0256	1.2092	1.3302	0.7292	1.0437	1.2184	1.4677	0.7257	1.6838	1.7268	1.6088	0.7703	1.2625	1.5938	1.6982

Table A.13: Average Performance Metrics for NASA Dataset

NASA Dataset		Average	
		RMSE	MAE
1D CNN	1 hour	0.5745	0.4149
	6 hours	1.6421	1.1679
	12 hours	2.4031	1.7834
	24 hours	3.0057	2.3050
LSTM	1 hour	0.4543	0.3091
	6 hours	1.5626	1.0957
	12 hours	2.4375	1.8155
	24 hours	3.0718	2.3156
WaveNet	1 hour	0.5085	0.3478
	6 hours	1.6052	1.1342
	12 hours	2.2661	1.6627
	24 hours	2.8894	2.1680
Transformer	1 hour	0.5106	0.3492
	6 hours	1.6235	1.1463
	12 hours	2.4259	1.7620
	24 hours	3.1401	2.3798
Swim4Ts	1 hour	1.0837	0.8222
	6 hours	1.8380	1.3594
	12 hours	2.6839	2.0586
	24 hours	3.1292	2.4540

Table A.14: Average Performance Metrics for KazHydroMet Dataset

KazHydroMet Dataset		Average	
		RMSE	MAE
1D CNN	1 hour	1.4581	1.0837
	6 hours	1.5937	1.1828
	12 hours	1.6921	1.2828
	24 hours	1.8414	1.3932
LSTM	1 hour	1.4412	1.0674
	6 hours	1.5241	1.1273
	12 hours	1.6737	1.2248
	24 hours	1.8661	1.3631
WaveNet	1 hour	1.5266	1.1331
	6 hours	1.7377	1.3104
	12 hours	1.8189	1.3560
	24 hours	1.8992	1.4254
Transformer	1 hour	1.2827	0.9390
	6 hours	1.4013	1.0189
	12 hours	1.5733	1.1413
	24 hours	1.7265	1.2668
Swim4Ts	1 hour	1.6446	1.2415
	6 hours	1.8123	1.3615
	12 hours	1.9331	1.4673
	24 hours	1.9674	1.5089

Table A.15: Average Performance Metrics for Yurt Dataset

Yurt Dataset		Average	
		RMSE	MAE
1D CNN	1 hour	0.9638	0.7369
	6 hours	1.7539	1.4095
	12 hours	2.0606	1.6467
	24 hours	1.8775	1.6503
LSTM	1 hour	0.9108	0.6977
	6 hours	1.3597	1.0256
	12 hours	1.6054	1.2092
	24 hours	1.5068	1.3302
WaveNet	1 hour	0.9527	0.7292
	6 hours	1.3711	1.0437
	12 hours	1.5767	1.2184
	24 hours	1.6561	1.4677
Transformer	1 hour	0.9491	0.7257
	6 hours	2.1238	1.6838
	12 hours	2.1800	1.7268
	24 hours	1.7585	1.6088
Swim4Ts	1 hour	1.0108	0.7703
	6 hours	1.6511	1.2625
	12 hours	2.0522	1.5938
	24 hours	2.0052	1.6982

Appendix B

Figures

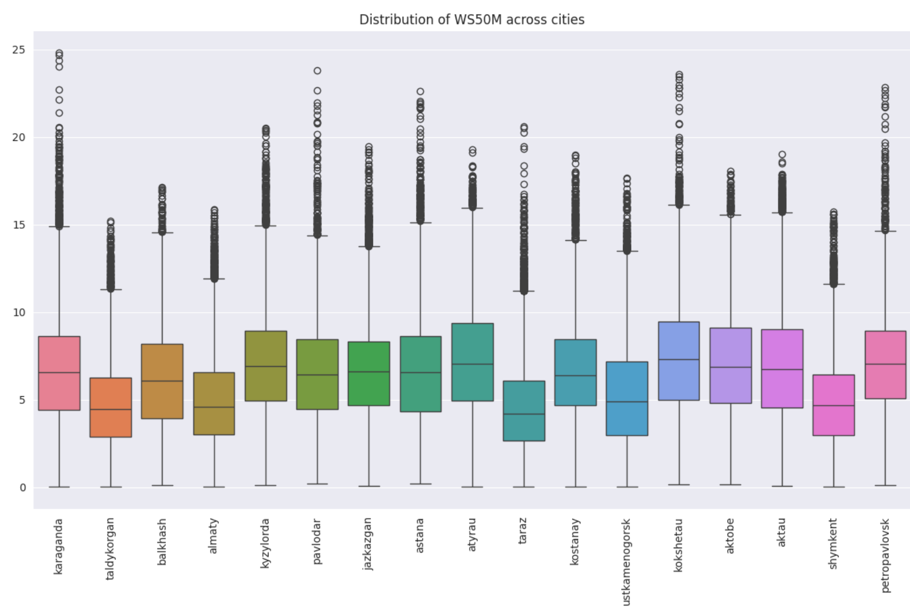

Figure B-1: Box-plot of the Wind Speed data from the NASA dataset across 17 cities in the Kazakhstan.

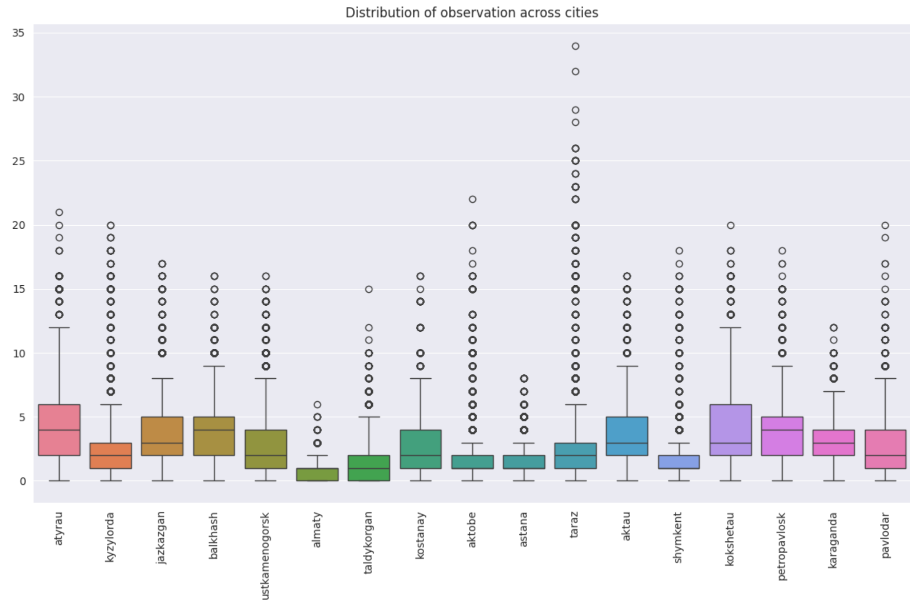

Figure B-2: Box-plot of the Wind Speed data from the KazHydroMet dataset across 17 cities in the Kazakhstan.

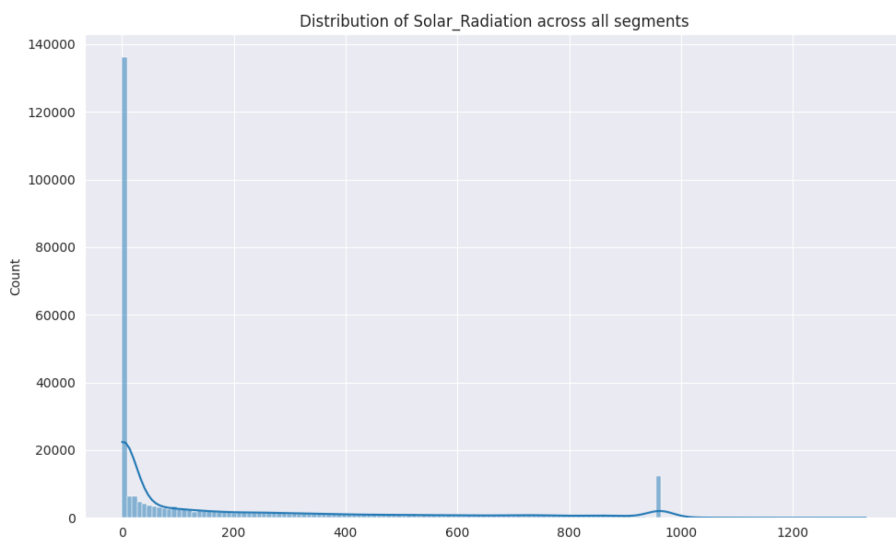

Figure B-3: Histogram of the Solar Radiation from the Yurt Dataset across all segments

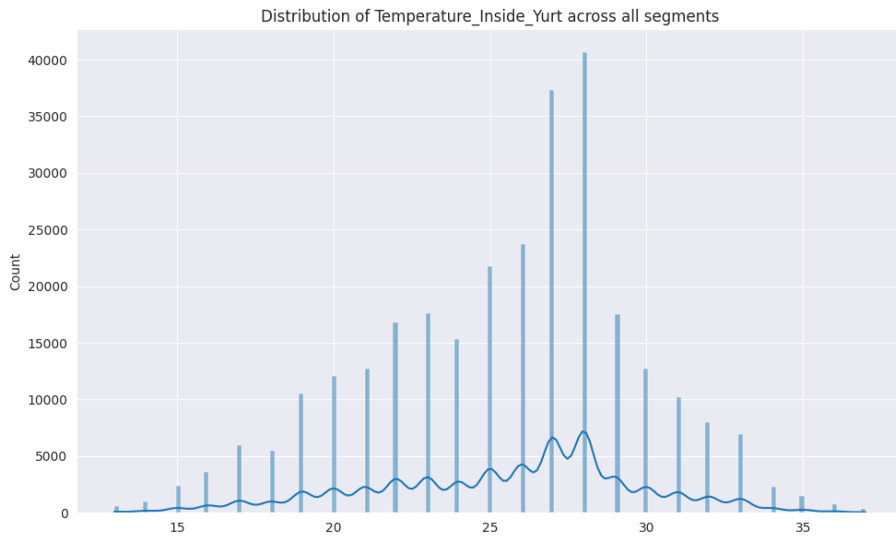

Figure B-4: Histogram of the Temperature Inside from the Yurt Dataset across all segments

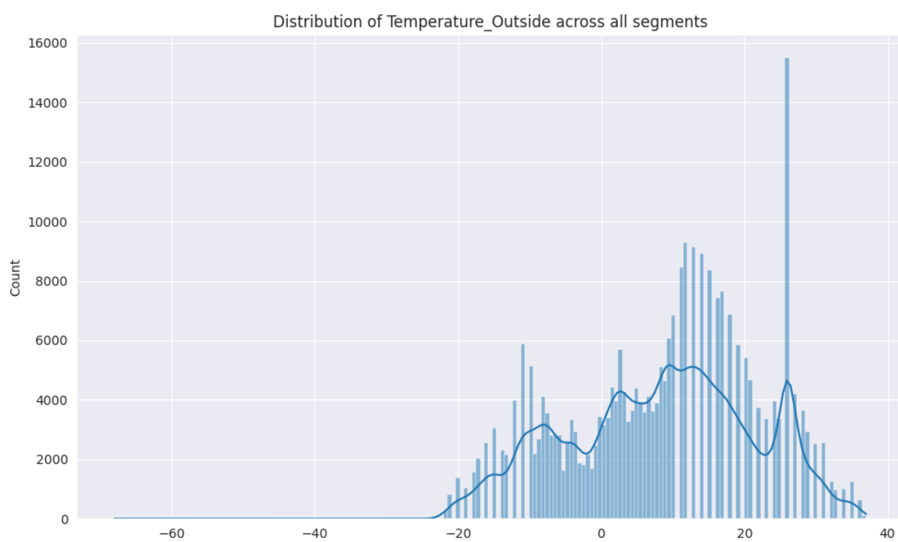

Figure B-5: Histogram of the Temperature Outside from the Yurt Dataset across all segments

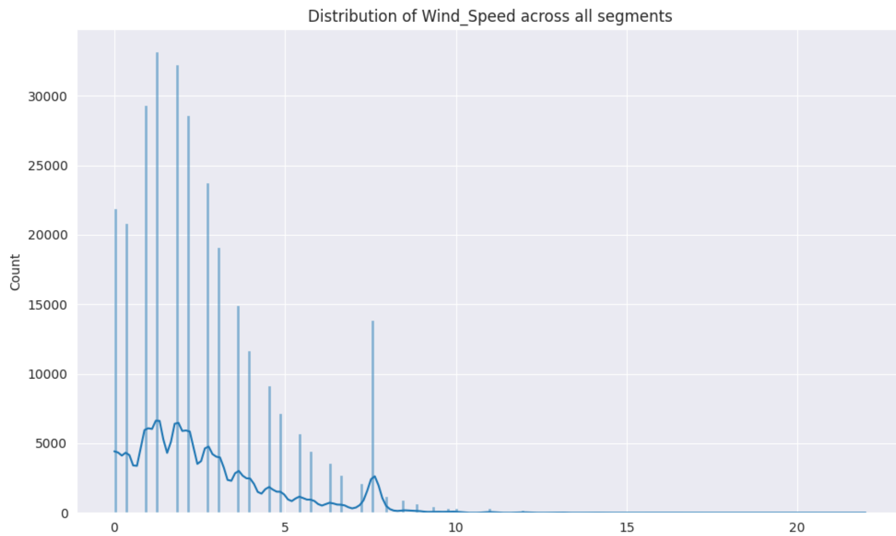

Figure B-6: Histogram of the Wind Speed from the Yurt Dataset across all segments

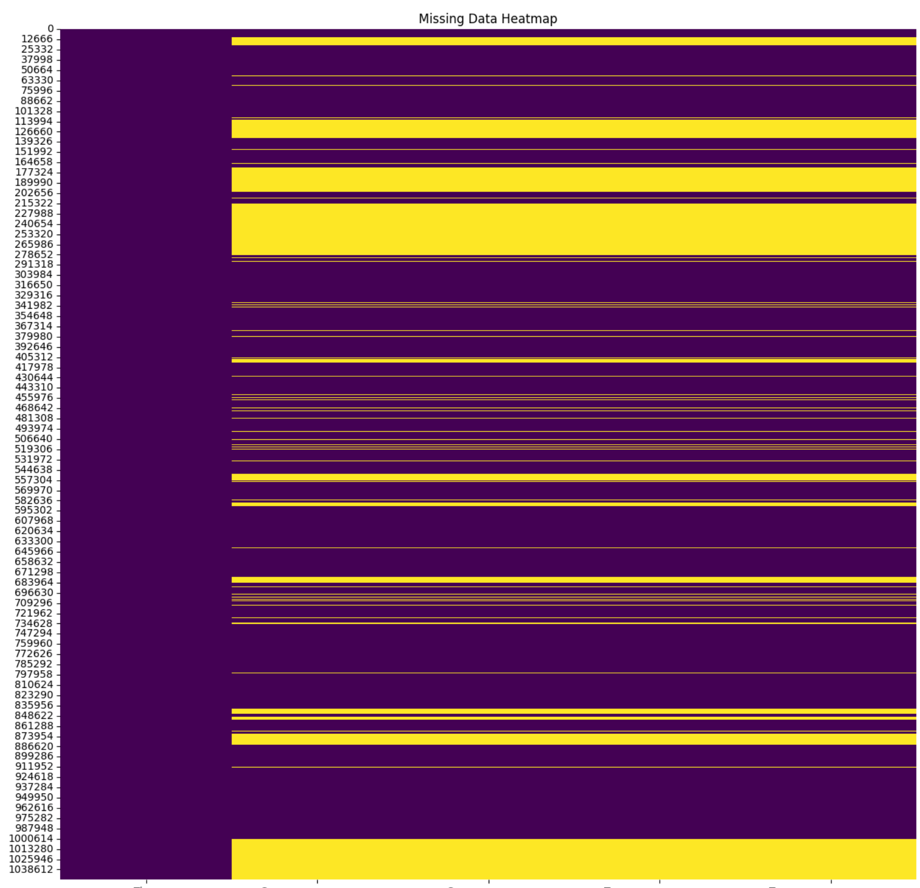

Figure B-7: Heatmap of missing values in the Yurt dataset from 2022 till 2023

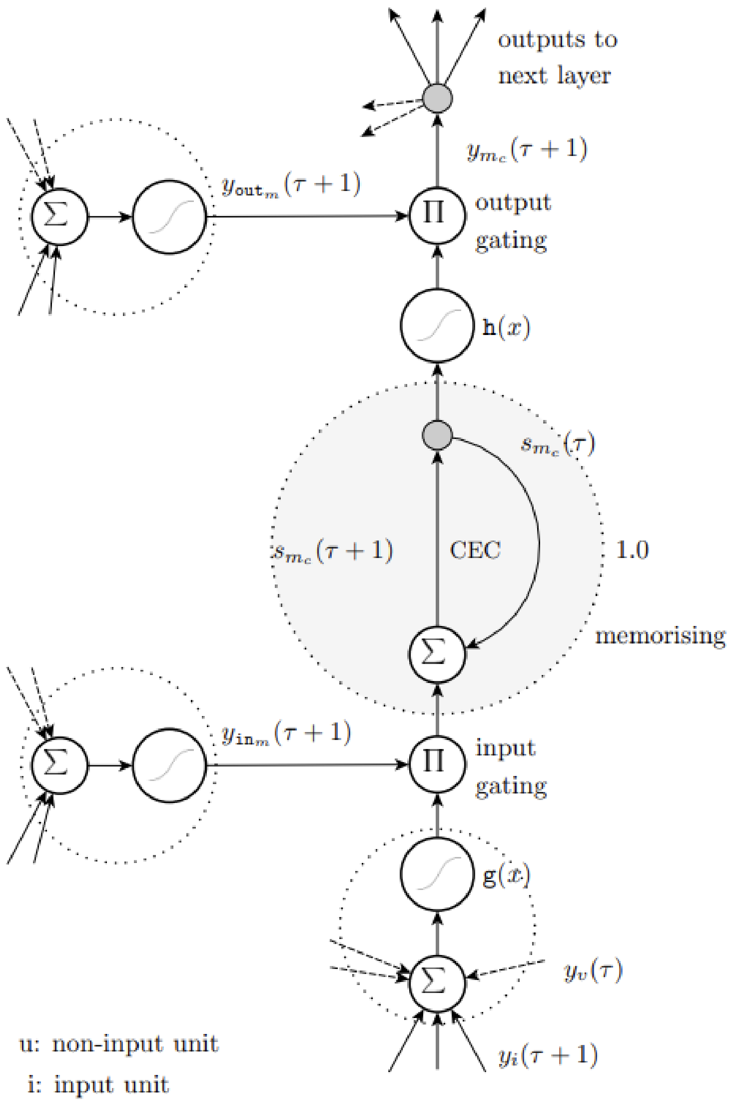

Figure B-8: Standart LSTM block architecture [4].

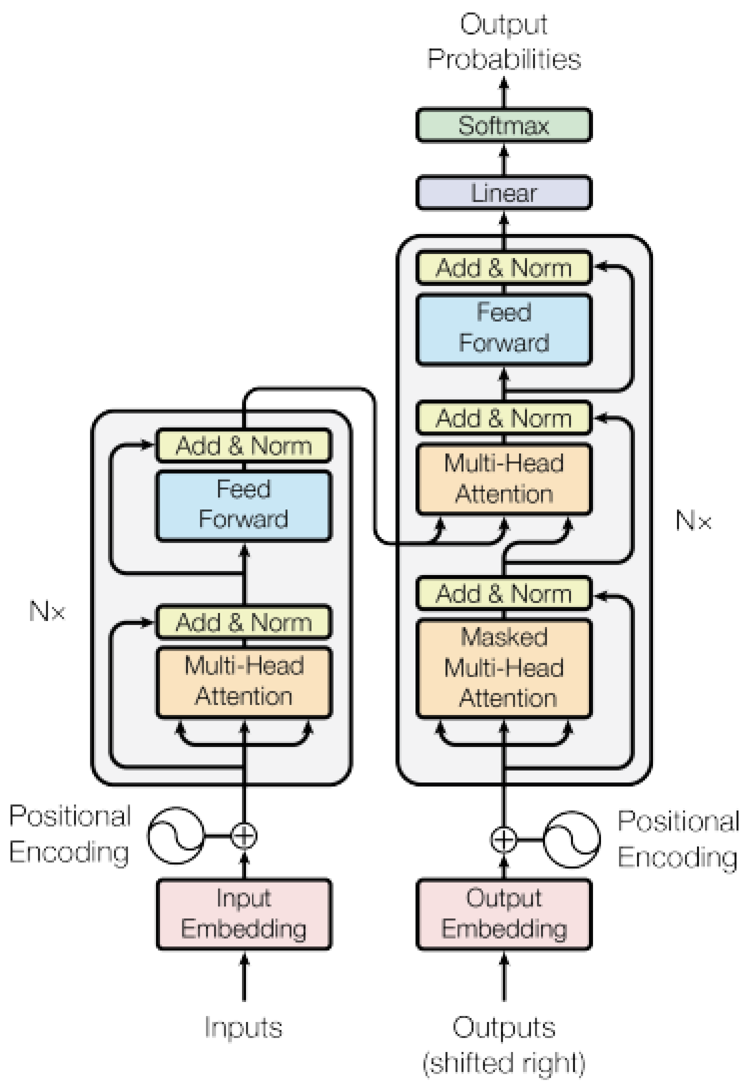

Figure B-9: Transformer architecture [5].

Bibliography

[1] Jaedong Kim, Seunghwan Oh, Heesoo Kim, and Woosung Choi. Tutorial on time series prediction using 1d-cnn and bilstm: A case example of peak electricity demand and system marginal price prediction. *Engineering Applications of Artificial Intelligence*, 126:106817, 2023.

[2] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio, 2016.

[3] Yunfei Du, Yin Wang, Ya Cong, Weihao Jiang, and Shiliang Pu. Long-term time series forecasting with vision transformer, 2024.

[4] Ralf C. Staudemeyer and Eric Rothstein Morris. Understanding LSTM - a tutorial into long short-term memory recurrent neural networks. *CoRR*, abs/1909.09586, 2019.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[6] Hamed H Pourasl and Vahid M Khojastehnezhad. Techno-economic analysis of wind energy potential in kazakhstan. *Proceedings of the Institution of Mechanical Engineers, Part A*, 235(6):1563–1576, 2021.

[7] Hakan Acikgoz and Deniz Korkmaz. Short-term offshore wind speed forecasting approach based on multi-stage decomposition and deep residual network with self-attention. *Engineering Applications of Artificial Intelligence*, 146:110313, 2025.

[8] Amirhossein Dolatabadi, Hussein Abdeltawab, and Yasser Abdel-Rady I. Mohamed. Hybrid deep learning-based model for wind speed forecasting based on dwpt and bidirectional lstm network. *IEEE Access*, 8:229219–229232, 2020.

[9] Lucas Hardy and Isla Finney. Leveraging state-of-the-art ai models to forecast wind power generation using deep learning. *Meteorological Applications*, 32, 03 2025.

[10] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR*, abs/2103.14030, 2021.

[11] Kazhydromet. Kazhydromet official website. `https://www.kazhydromet.kz/en`, 2025. Accessed: 2025-03-20.

[12] NASA POWER. Nasa power data access viewer. `https://power.larc.nasa.gov/data-access-viewer/`, 2025. Accessed: 2025-03-20.

[13] Schalk van der Merwe, Armand du Plessis, and Arnold Rix. Wind power forecasting: A review. 11 2023.

[14] Zongxu Liu, Hui Guo, Yingshuai Zhang, and Zongliang Zuo. A comprehensive review of wind power prediction based on machine learning: Models, applications, and challenges. *Energies*, 18(2), 2025.

[15] Shahram Hanifi, Xiaolei Liu, Zi Lin, and Saeid Lotfian. A critical review of wind power forecasting methods—past, present and future. *Energies*, 13(15), 2020.

[16] Yun Wang, Runmin Zou, Fang Liu, Lingjun Zhang, and Qianyi Liu. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy*, 304:117766, 12 2021.

[17] Wen-Chang Tsai, Chih-Ming Hong, Chia-Sheng Tu, Whei-Min Lin, and Chiung-Hsing Chen. A review of modern wind power generation forecasting technologies. *Sustainability*, 15(14), 2023.

[18] Zongxu Liu, Hui Guo, Yingshuai Zhang, and Zongliang Zuo. A comprehensive review of wind power prediction based on machine learning: Models, applications, and challenges. *Energies*, 18(2), 2025.

[19] Wen-Chang Tsai, Chih-Ming Hong, Whei-Min Lin, Chia-Sheng Tu, and Chiung-Hsing Chen. A review of modern wind power generation forecasting technologies, 04 2023.

[20] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021.

[21] Kai Qu, Gangquan Si, Zihan Shan, XiangGuang Kong, and Xin Yang. Short-term forecasting for multiple wind farms based on transformer model. *Energy Reports*, 8:483–490, 2022. ICPE 2021 - The 2nd International Conference on Power Engineering.

[22] Huijuan Wu, Keqilao Meng, Daoerji Fan, Zhanqiang Zhang, and Qing Liu. Multistep short-term wind speed forecasting using transformer. *Energy*, 261:125231, 2022.

[23] Erick Giovani Sperandio Nascimento, Talison A.C. de Melo, and Davidson M. Moreira. A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy. *Energy*, 278:127678, 2023.

[24] Yun Wang, Tuo Chen, Shengchao Zhou, Fan Zhang, Ruming Zou, and Qinghua Hu. An improved wavenet network for multi-step-ahead wind energy forecasting. *Energy Conversion and Management*, 278:116709, 2023.

[25] Shubhi Harbola and Volker Coors. One dimensional convolutional neural network architectures for wind prediction. *Energy Conversion and Management*, 195:70–75, 2019.

[26] Hui Liu, Xi wei Mi, and Yan fei Li. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and elman neural network. *Energy Conversion and Management*, 156:498–514, 2018.

[27] Sixian Yue, Zhili Zhao, Tianyou Lai, and Jin Zhang. A modularity-enhanced echo state network for nonlinear wind energy predicting. *Energies*, 18(7):1858, 2025.

[28] Xiangqian Li, Keke Li, Siqi Shen, and Tian Yaxin. Exploring time series models for wind speed forecasting: A comparative analysis. *Energies*, 16:7785, 11 2023.

[29] Xiaoli Ren, Xiaoyong Li, Kaijun Ren, Junqiang Song, Zichen Xu, Kefeng Deng, and Xiang Wang. Deep learning-based weather prediction: A survey. *Big Data Research*, 23:100178, 12 2020.

[30] Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200209, February 2021.

[31] Xuanfang Da, Dong Ye, Yanbo Shen, Peng Cheng, Jinfeng Yao, and Dan Wang. A novel hybrid method for multi-step short-term 70 m wind speed prediction based on modal reconstruction and stl-vmd-bilstm. *Atmosphere*, 15(8), 2024.