18th International Conference on Environmental Science and Technology

Athens, Greece, 30 August to 2 September 2023

# SWIRL: Statistical downscaling for Wind Pattern Reconstruction using Machine Learning

**Skianis K.[1,*], Giannopoulos A.[1], Spantideas S.[1], Hatzaki M.[2], Karditsa A.[1], Trakadas P.[1]**

[1] Department of Ports Management and Shipping, National and Kapodistrian University of Athens

[2] Department of Geology and Geoenvironment, National and Kapodistrian University of Athens

*corresponding author:
e-mail: kskianis@pms.uoa.gr

**Abstract** Ports are critical infrastructures for global supply chains, crucial hubs and strategic to future trade. However, they are particularly exposed to Climate Change (CC) impacts, estimated to have broad implications on economy and human welfare. Therefore, a timely introduction of adaptation measures addressing CC impacts on ports becomes a major priority and can be proactive if based on projected climate. Yet, this challenge requires high spatial resolution timeseries for the present and the projected climate which are frequently missing. Moreover, employed downscaling procedures are not always skillful, particularly for extremely complex wind fields. The scope of this study is the development of reliable high-resolution wind speed/direction timeseries through Machine Learning (ML) techniques application. The employed ML regression schemes exploit ECMWF-ERA5 Reanalysis data as input training dataset (10931 instances) for Heraclion port area (Crete-Greece), containing 1 site of interest and 4 peripheral (period 1975-2004). Analytical simulations were conducted towards evaluating the regression accuracy on test data in terms of the Mean Absolute Error (MAE). Study outcomes revealed that ML techniques can efficiently reconstruct wind speed/direction timeseries, contributing to the wind downscaling and reconstruction problem, capable of supporting stakeholders needs on port scale regarding CC adaptation.

**Keywords: wind speed, wind direction, wind pattern, machine learning, statistical downscaling**

## 1. Introduction

Wind patterns are an essential component of weather and climate systems, impacting many aspects of human life, including agriculture, transportation, energy production and trade. Ports, which are key elements of trade and commerce are particularly vulnerable to the effects of Climate Change (CC). Having access to data and being able to predict the wind is critical for ports (Solari et al., 2012), as they serve 80% of world trade. There is an evident need for data, especially in ports and their surrounding areas where no meteorological stations are available (Rodriguez et al., 2017). However, obtaining high-resolution wind data is often challenging, as it requires expensive equipment and significant resources and wind downscaling (measure and direction) is one of the greatest challenges in climatology (Pryor and Hahmann, 2019).


*Figure 1: the datapoints of interest.*

SWIRL provides a solution to this problem by leveraging machine learning algorithms to reconstruct wind data by exploiting existing data of nearby stations. This approach involves training machine learning models on high-resolution wind data and then using it to predict wind patterns at a lower resolution. We test our approach on a real dataset, where we want to reconstruct, in a supervised way, wind speed and direction data by using information from neighboring meteorological stations. Figure 1 illustrates the point of interest (EMY-yellow pin) which we want to reconstruct with the help of the surrounding 4 green points (era1-era4). The yellow pin is located near the Greek port of Heraklion, Crete, and wind data is of paramount importance for the specific infrastructure.

In this work, we examined the case of reconstructing data for a port of interest as ports constitute crucial infrastructure. The ultimate goal of SWIRL is to exploit data that are available in a given area and then use the model's weights to reconstruct data in a point of interest of a geographically/morphologically similar area, in the case that the distributions of the new surrounding points follow the ones of the previous area. The potential applications of SWIRL are extensive, including improving weather forecasting accuracy, enabling more effective energy production, and supporting climate change research. In this paper, we provide a detailed description of the SWIRL approach, its implementation, and its performance in reconstructing wind speed and direction data.

## 2. Related work

Rodríguez et al. (2017) presented a hybrid methodology using a compact genetic algorithm with an artificial neural network for wind speed time series reconstruction. Later, Lazoglou et al. (2019) introduced a novel statistical method combining triangular irregular networks and links to simulate extreme maximum and minimum temperatures. Hu et al. (2020): Deterministic and probabilistic wind speed prediction with denoising-reconstruction strategy and quantile regression-based algorithm. Jing et al. (2022) proposed an improved environment coder (ICE) network with multiple one-dimensional convolutional layers (CNN) for wind speed data reconstruction. While the aforementioned methods used novel approaches and deep learning architectures, they did not use data of nearby stations to exploit existing knowledge and reconstruct data in a specific point of interest.

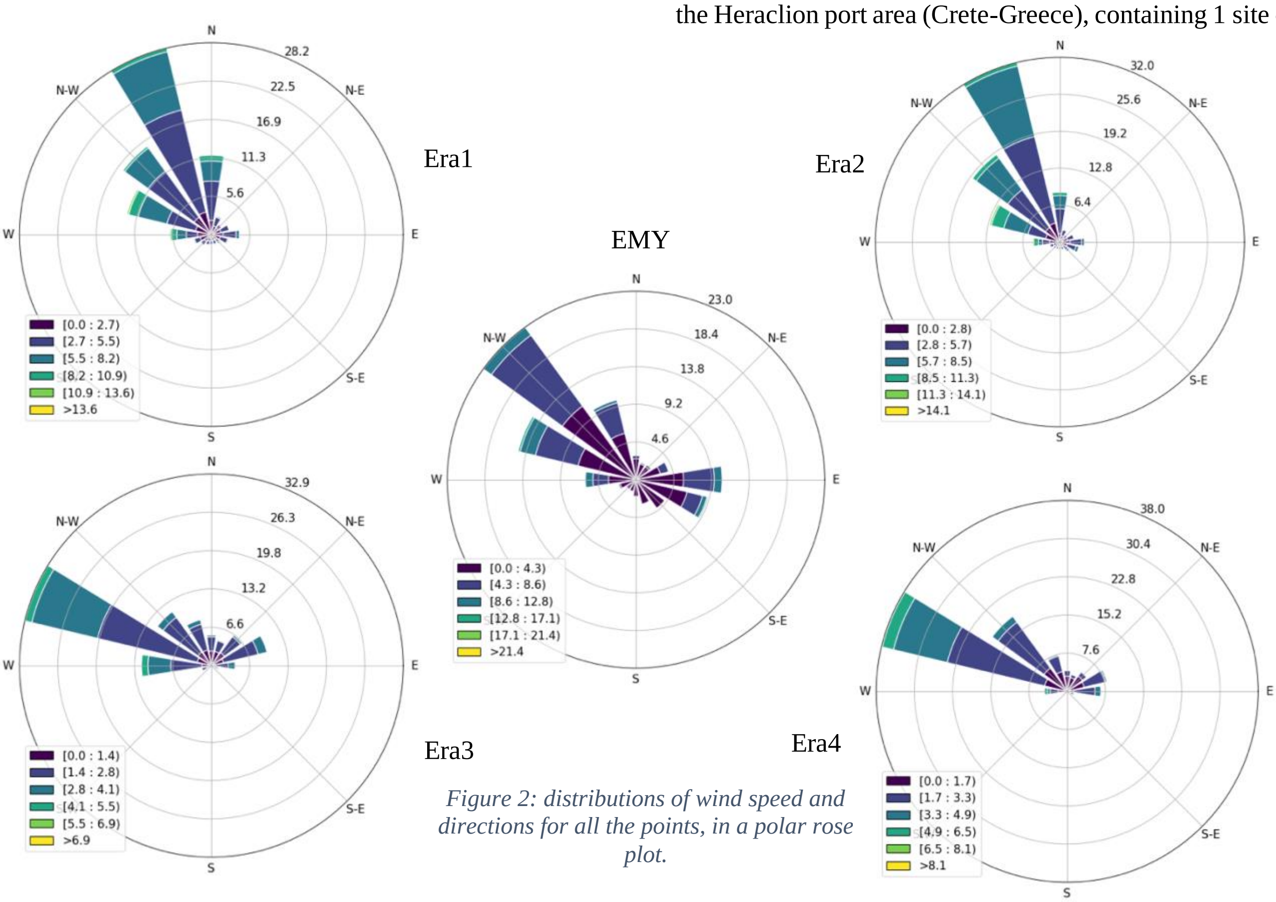


*Figure 2: distributions of wind speed and directions for all the points, in a polar rose plot.*

## 3. Dataset

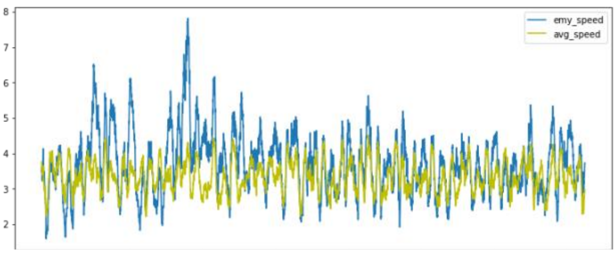


*Figure 3: average speed of the 4 surrounding points compared to the wind speed of middle point.*

The dataset consists of daily zonal and meridional surface wind components from the ECMWF ERA5 Reanalysis database with a spatial resolution of 0.25x0.25 degrees for the period 1975-2004. More specifically we have daily values of wind speed and direction (10931 instances) for the Heraclion port area (Crete-Greece), containing 1 site of interest and 4 peripheral sites. Speed was measured initially in knots and direction was measured in terms of degrees (0-360$^0$).

Figure 2 shows the distributions of the 4 neighboring points (Era points). In terms of direction we quickly observe that the two north points have a similar behavior of North to North-West, the two south points follow a West to North-West behavior while the EMY point indicates mostly North-West. Figure 3 presents the average speed of the 4 surrounding points, showing that a simple averaging method could not be effectively used for reconstruction.

## 4. Experiments

The purpose of our experiments is to examine reconstruction of air data (speed and direction) using data from geographically close points. Next, we present all steps of the SWIRL approach, including preprocessing, analysis, selecting machine learning models for reconstruction and finally metrics of evaluation. For preprocessing, we converted speed from knots to m/s, and decomposed direction of the surrounding points to sine and cosine. The machine learning models used to reconstruct the wind timeseries come from the areas of forecasting and regression. For forecasting, we only use past values and not data from surrounding points, while for regression the 4 values of the surroundings stations as features, and the values of the middle point (EMY) as the target.

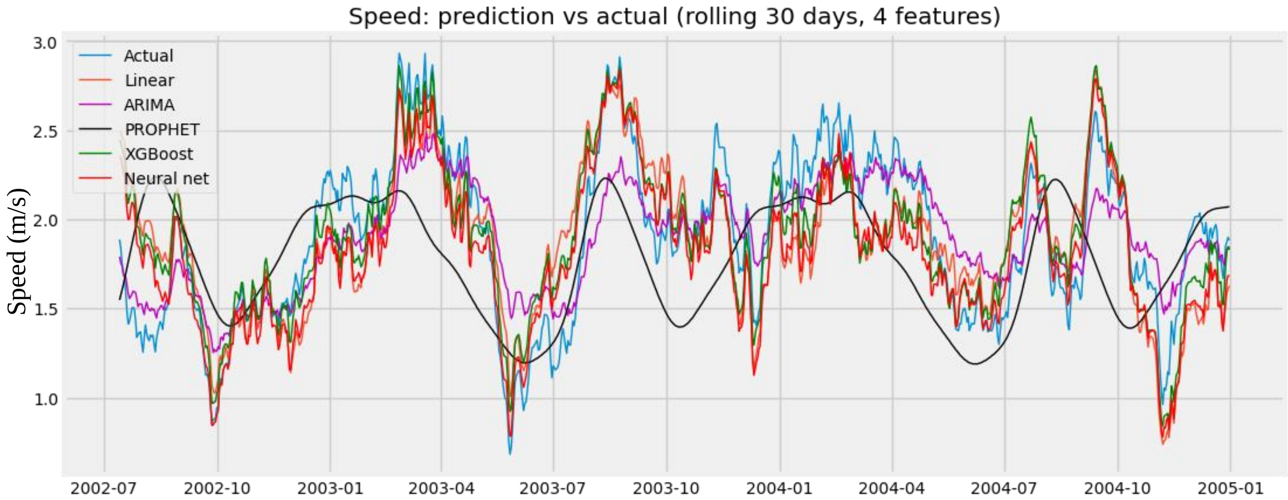


*Figure 4: speed prediction vs ground truth, averaged by month.*

We use a sliding window of 7 days to look back, and split the dataset into 80% for training and 20% for testing-reconstruction. The evaluation metric used was the Mean Absolute Error (MAE). Multiple ML-based models were deployed and compared for selecting the best ML regressor, utilizing deterministic and simplistic benchmarks. After extensive experiments we selected 2 forecasting and 3 regression models, that provided the best results in the testing subset.

The first forecasting model to be used is ARIMA, which stands for Autoregressive integrated moving average, is a statistical analysis model that uses values based on previous values. Next, the second forecasting method to be exploited is PROPHET[1], a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data.

For the first regression model we examine linear regression, a method for modeling the connection between a scalar response and one or more explanatory factors using a linear approach. Relationships are modeled using linear predictor functions whose unknown model parameters are derived from data.

*Table 1: results on wind speed prediction.Bold indicates best method with lower Mean Absolute Error (MAE).*

| Model | MAE (4 feats) | MAE (4 feats plus dir) |
|---|---|---|
| Average | 1.4931 | - |
| ARIMA | 1.0107 | - |
| PROPHET | 1.1657 | - |
| Linear reg. | 0.5733 | 0.5161 |
| XGBoost | 0.4653 | 0.3966 |
| Neural | **0.4558** | **0.3831** |

XGBoost (eXtreme Gradient Boosting) is a free and open-source software library that provides a regularizing gradient boosting framework for a variety of computer languages, including Python. XGBoost operates as Newton-Raphson in function space, and a second order Taylor approximation is employed in the loss function to create the connection to the Newton Raphson approach. An artificial neural network is made up of artificial neurons or nodes, used to solve artificial intelligence (AI) problems.

*Table 2: results on wind direction prediction. Bold indicates best method with lower Mean Absolute Error (MAE).*

| Model | MAE (4 feats) |
|---|---|
| Average | 63.80 |
| Linear reg. | 62.25 |
| XGBoost | 51.77 |
| Neural | **50.54** |

When processing samples that each have a known "input" and "output," neural networks learn by creating probability-weighted connections between the two that are then stored inside the net's data structure. To train it, we compare the output against the desired output. The network then modifies its weighted associations using this error value and a learning strategy. Here, we build a model of two dense layers followed by Relu activation functions and one final dense layer for the outcome. The loss to be used is Mean Absolute Error (MAE) and the optimizer is Adam (Kingma and Ba, 2014).

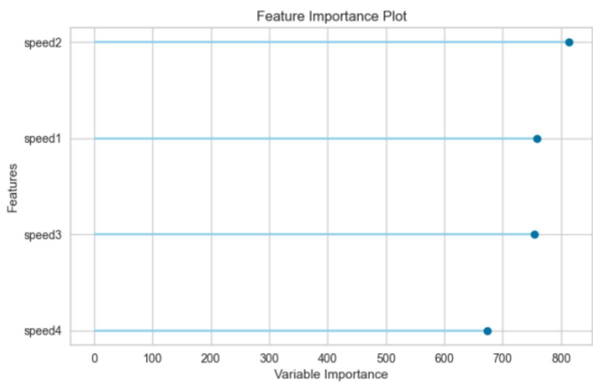


*Figure 5:speed feature importance for the 4 nearby points.*

Table 1 presents the results on wind speed prediction with all machine learning models. First averaging the

[1] https://facebook.github.io/prophet/

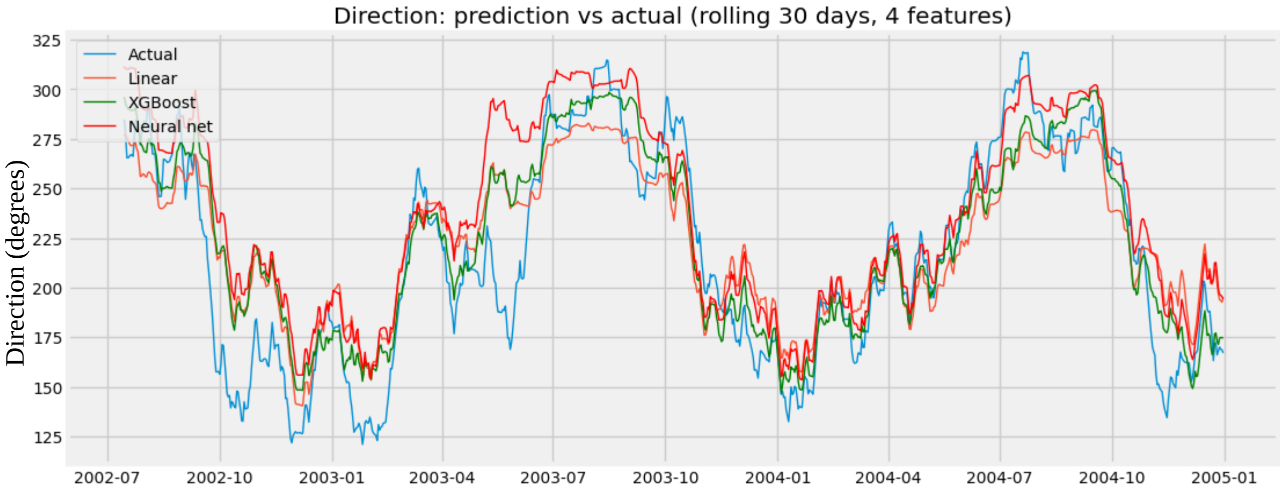


*Figure 6: direction prediction vs ground truth, averaged by month.*

speeds of the 4 surrounding points is used as baseline approach. Next, the forecasting techniques ARIMA and PROPHET prove to be more effective than the averaging approach. The neural network approach was the best approach with the lower error, beating linear regression and XGBoost. We also show the effective use of direction as extra features, reducing the error. Figure 4 shows the reconstructed wind speed timeseries by all methods compared to the actual ground truth (blue line), averaged by month so that the different results are visible. Table 2 shows the results on predicting the direction of wind with averaging, linear regression, XGBoost and a neural network, with the neural network being again the best because of its ability to capture non-linear and complex interactions. In order to further analyze how each surrounding point contributes to the reconstruction, Figure 5 indicates the feature importance for the speed variables, showing that north points contribute more. Finally, Figure 6 presents the reconstructed wind direction timeseries by all models compared to the actual ground truth (blue line), again averaged by month. We clearly see that predicting the wind direction is harden than predicting the wind speed.

## Conclusion and Future Work

Our contributions are summarized in the following points: a) analysis and study of wind time series in speed and direction, b) successful integration of machine learning models for time series, c) use of neighboring points to improve the reconstruction. To summarize, SWIRL was essentially designed to exploit data available in a given area and then use the model's weights to reconstruct data in a point of interest of a geographically similar area. The case of reconstructing data for a port of interest was examined as ports constitute crucial infrastructure. Morevover, as Greece's economy is highly dependent on the smooth operation of ports. Being able to reconstruct wind data for areas that did not previously have meteorological stations could help significantly towards improved weather prediction.

As future work we plan to integrate SWIRL to a European project entitled "AdaptPorts" which is focused on Strategic Action for the Mitigation and Adaptation of Ports to Climate Change. A next step is to exploit more complex machine learning models, for example using statistical models with distribution distances or new deep learning techniques like Generative Adversarial Networks (GANs) that have proven to be very effective in the case of timeseries reconstruction.

## References


Chang, Wen-Yeau. "A literature review of wind forecasting methods." *Journal of Power and Energy Engineering* 2.04 (2014): 161.

Hu, Jianming, et al. "Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm." *Renewable Energy* 162 (2020): 1208-1226.

Jing, Bo, et al. "Missing wind speed data reconstruction with improved context encoder network." *Energy Reports* 8 (2022): 3386-3394.

Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).

Lazoglou, Georgia, Benedikt Gräler, and Christina Anagnostopoulou. "Simulation of extreme temperatures using a new method: TIN-copula." *International Journal of Climatology* 39.13 (2019): 5201-5214.

Pryor, S. C., and Andrea N. Hahmann. "Downscaling wind." *Oxford Research Encyclopedia of Climate Science*. 2019.

Rodriguez, H., et al. "Wind speed time series reconstruction using a hybrid neural genetic approach." *IOP Conference series: earth and environmental science*. Vol. 93. No. 1. IOP Publishing, 2017.

Solari, Giovanni, et al. "The wind forecast for safety management of port areas." *Journal of Wind Engineering and Industrial Aerodynamics* 104 (2012): 266-277.

Taylor, Sean J., and Benjamin Letham. "Forecasting at scale." *The American Statistician* 72.1 (2018): 37-45.

Wu, Yuan-Kang, and Jing-Shan Hong. "A literature review of wind forecasting technology in the world." *2007 IEEE Lausanne Power Tech* (2007): 504-509.