A multifidelity framework for wind speed data

Pietro Colombo[1], Claire Miller [1], Ruth O'Donnell[1], Xiaochen Yang[1]

[1] University of Glasgow, UK

E-mail for correspondence: pietro.colombo@glasgow.ac.uk

Abstract: Monitoring wind speed is essential to develop offshore wind farms. However, recorded wind data often lack the necessary accuracy for understanding the profitability of the wind farm, and even when they exist, they are scarce in time or space. Intuitively, using multiple data sources could balance the trade-off between scarcity and accuracy. A multi-fidelity framework in the form of the autoregressive Gaussian process is introduced to analyze wind speed reanalysis data fusing datasets of different reliability and resolution to provide a more accurate wind speed data product.

Keywords: Multifidelity; Gaussian process; Wind speed.

1 Introduction

Offshore wind speed data are obtained through different means such as in-situ field sampling using anemometric or Lidar technologies, or processed satellite retrievals. The former provides high-fidelity (high quality) and high-resolution measurements but is limited in temporal and spatial coverage, while the latter offers larger coverage but with low-fidelity (low quality) and low-resolution. Data fusion of two products can, in principle, provide a more informative data stream.
The development of a series of offshore wind farms on the Italian Adriatic coast is our motivational study case. The project known as Agnes (Adriatic green network of energy sources) aims to build a hub for renewable energy. For the wind speed data, the companies involved in the project rely on two main data sources:

1. The ERA5 reanalysis data [ERA5 Data], which contains hourly wind speed measurements from 1979 until the present.

This paper was published as a part of the proceedings of the 37th International Workshop on Statistical Modelling (IWSM), Dortmund, Germany, 16–21 July 2023.

2. The wind climatology obtained through two Lidar installations. These measurements tend to be more reliable than those of ERA; however, they come as point samples, covering a minimal spatial surface.

We developed a simulation study using ERA5 reanalysis data (from the Agnes location), from which we derived two datasets, one of high-fidelity (HF), closer to the true wind speed, but with low temporal sampling rate, the other of low-fidelity (LF) but with high temporal sampling rate. This paper evaluates the performance of an autoregressive Gaussian process (ARGP) [Le Gratiet L. & Garnier J.(2014)] for data fusion of multi-fidelity data. Through the multi-fidelity framework, we aim to return predictions that are more accurate and abundant than those based only on a single data source.

2 Experimental Design

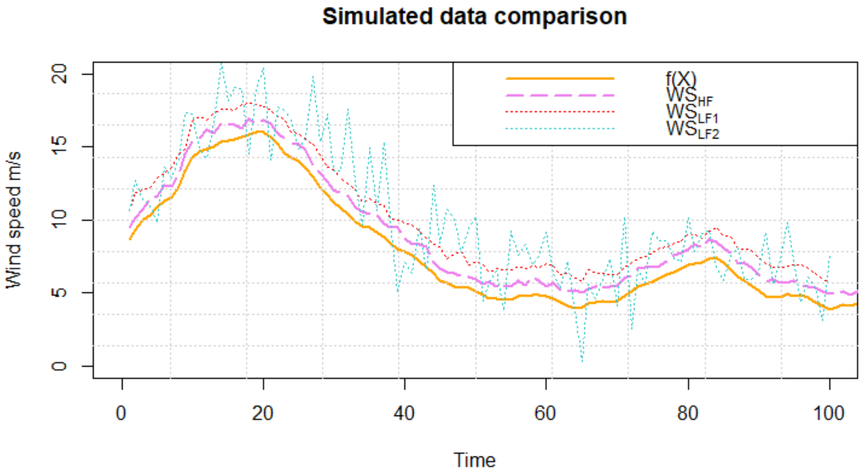
We have simulated two data sources (time-series) that resemble wind speed measurements such that:

$$ws_{HF} = f(x) + e_{HF}(x) \tag{1}$$

$$ws_{LF} = f(x) + e_{LF}(x). \tag{2}$$

The ws_{HF} represents the high-fidelity measurements, therefore closer to $f(x)$, the true wind speed at the index location x, while ws_{LF} is a low-fidelity measurement. The time-series are distinguished by the normally distributed corruptions e_{HF} and e_{LF}, with the low-fidelity corruptions ($e_{LF1} \sim N(2, 0.2)$ and $e_{LF2} \sim N(2, 1)$) being roughly double that of the high-fidelity corruption ($e_{HF} \sim N(1, 0.2)$). HF data are typically scarce, therefore we performed an additional sampling of them of size $N1 < N$. Starting from the ERA5 reanalysis wind speed data, we proceed with a series of decompositions to extract the deterministic part of these data: $f(x)$, which we assume to be composed of a long term trend, a seasonal pattern and potentially other cyclical components. Given a roughly normal remainder, we can generate different ws_{HF} and ws_{LF} adding normally distributed errors. Figure 1 depicts an example of data constructed with such an approach. Given such a design, we compared the model performances of ARGP with two mono-fidelity models: the quantile gradient-boosted regression tree (QGBRT) [Kriegler B. & Berk R. (2010)], a model often used in wind forecasting that provides the prediction for all quantiles of a distribution (hence a deeper understanding of the uncertainty), and a standard GP. We controlled for different $N1$ sample sizes and used as a performance metric the mean absolute deviation (MAE) of the residuals $r = f(x) - P_i$, where P_i is the i^{th} model prediction. The experimental comparison has $N = 850$, 100 replications of randomly drawn errors e_{LF} and e_{HF} and the sub-sample $N1$ index position.

FIGURE 1: Comparison of the four generated signals: in orange $f(x)$ the assumed true wind speed, in violet ws_{HF} the high-fidelity time-series, in dark orange a low noise low-fidelity time-series ws_{LF1} and in turquoise a noisy version of the low-fidelity data ws_{LF2}.



3 Models

3.1 Multifidelity:ARGP

In the ARGP model, the high-fidelity data are modelled as a scaled sum of the lower-fidelity data:

$$GP_{HF}(x) = \rho GP_{LF}(x) + \epsilon(x), \tag{3}$$

where $GP_{HF}(x)$ is a Gaussian process modelling the HF data, $GP_{LF}(x)$ a Gaussian process modelling the LF data, ρ is the degree of correlation between the HF and LF data and $\epsilon(x) \sim GP(\mu_\epsilon, \Sigma_\epsilon)$ is an independent Gaussian process denoting the error structure between HF and LF data. Our simulation design presents a nested structure $D_{HF} \subset D_{LF}$; therefore, we can use the recursive formulation of the model proposed by [Le Gratiet L. & Garnier J.(2014)], which guarantee an efficient maximum likelihood inference.

3.2 Monofidelity: GP and QGBRT

In opposition to ARGP, we tested a standard Gaussian process (GP) fitted only with LF and HF separately, denoted by the notation GP_{HF}, GP_{LF} and a QGBRT in which a quantile loss function is combined with a gradient-boosted regression tree, also fitted using only one dataset and denoted by $QGBRT_{LF}$ and $QGBRT_{HF}$.

4 Results and Discussion

By comparing the models, with x being a time index, for different $N1$ high fidelity sample sizes, and low-fidelity noise setting $e_{LF1} \sim N(2, 0.2)$, we obtained the results in Table 1. ARGP outperformed the other models for small $N1$ sample sizes, while for $N1 > 160$ its performance was equivalent to those of GP_{HF} and $QGBRT_{HF}$. It also appears there is no notable advantage with highly noisy data. For our simulation design, with $N1$=32, the estimates from the ARGP which combined the low and high-fidelity data were, on average, 1 m/s closer to the $f(x)$, which consists of a 16% improvement compared to the unprocessed information. The multi-fidelity framework has been successfully applied to multiple environmental applications; however, its application to a wind case study is new. This work has illustrated that potentially modelling wind speed data with the multi-fidelity framework is appropriate. To further improve the methodology, three directions of future development have been identified: expansion to include the spatial dimension, exploration of non-linear methodologies for noisy data, and integration of techniques to address data skewness.

TABLE 1: MAE summary from 100 replications for a simulationw with different $N1$ high-fidelity samples size, with low fidelity data error structure equal to $e_{LF} \sim N(2, 0.2)$; The table contains the performance of 5 models: two mono-fidelity GP, two mono-fidelity QGBRT and a multi-fidelity ARGP. In the parenthesis, the MAE standard deviation.

MODELS	$N1$=32(sd)	$N1$=96(sd)	$N1$=160(sd)
GP$_{LF}(Time)$	0.54(0.018)	0.54(0.018)	0.54(0.018)
GP$_{HF}(Time)$	0.43(0.048)	0.32(0.011)	0.30(0.008)
ARGP$(Time)$	0.29(0.047)	0.29(0.060)	0.29(0.015)
QGBRT$_{LF}(Time)$	0.55(0.020)	0.55(0.020)	0.55(0.020)
QGBRT$_{HF}(Time)$	0.52(0.005)	0.39(0.021)	0.34(0.016)

References

ERA5 Data https://cds.climate.copernicus.eu/cdsapp!/dataset/reanalysis-era5-pressure-levels?tab=overview.

Kriegler, B., Berk, R. (2010) Small area estimation of the homeless in Los Angeles: An application of cost-sensitive stochastic gradient boosting. *Ann. Appl. Stat* 4(3): 1234-1255

Le Gratiet, L., Garnier, J. (2014) Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5).