

PAPER • OPEN ACCESS

Wind field reconstruction from lidar measurements at high-frequency using machine learning

To cite this article: Clym Stock-Williams *et al* 2018 *J. Phys.: Conf. Ser.* **1102** 012003

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

Wind field reconstruction from lidar measurements at high-frequency using machine learning

Clym Stock-Williams¹, Paul Mazoyer² and Sébastien Combrexelle³

¹ TNO, Westerduinweg 3, 1755 LE Petten, The Netherlands

² Leosphere, 14-16 rue Jean Rostand, 91400 Orsay, France

Clym.Stock-Williams@tno.nl

Abstract. The increase in wind turbine size during recent years has led to a situation where the wind speed measured at hub height is no longer a sufficient representation of the power available to be extracted by a turbine. Light Detection and Ranging (“Lidar”) wind sensing devices can measure wind speeds in a volume in front of the sensor, offering affordable and improved estimation of wind turbine power performance and loading. However, a Lidar can only measure the component of the wind velocity in the direction of the beam. This makes reconstruction of the wind velocity field from raw Lidar measurements a challenge, and current methods contain assumptions which cause inaccuracies in complex terrain and in wind turbine wakes. In addition, data availability is reduced in specific weather conditions, such as very cold climates and dense fog. A novel Machine Learning method is developed here, based on Gaussian Process regression, to remove these assumptions when producing full 3D wind fields from Lidar measurements. This approach is naturally robust to overfitting and predicts uncertainty derived from data density and machine error, without needing physical information about the local terrain. Initial validation by comparison of a Windcube Lidar with a meteorological mast, on a site with simple conditions, shows excellent performance which meets the threshold for commercial use. The method also infers data during measurement gaps, offering the potential for 100% data availability. Instantaneous 2D wake measurements in a plane can be made, which when validated and further developed will offer a breakthrough in understanding of wind farm interactions and turbine loading.

1. Introduction

The increase in size of wind turbines over the recent years has led to a situation where the wind speed measured at hub height is no longer a sufficient representation of the power available to be extracted by a turbine. The new version of IEC standard 61400-12-1 therefore allows for a new definition: Rotor Equivalent Wind Speed, which considers the variation of wind speed and angle across the rotor area.

Such a definition, however, requires data to calculate it. Light Detection and Ranging (“Lidar”) wind sensing devices have emerged in recent years, which use lasers to measure wind velocity over a volume (a 3D “wind field”) in front of the device. Ground-based Lidar measurements are approved in the new standard for power performance testing, although still when accompanied by a meteorological mast.

This advantage over traditional point-measurement anemometers, along with their portability, have enabled not only more accurate power performance measurements, but also several innovations which could reduce Cost of Energy (“CoE”) significantly:



- *Yaw misalignment correction.* This has already demonstrated up to 5% improvement in power output.
- *Site suitability pre-construction studies,* to select appropriate turbines for the site, considering the wind speeds over the full area of the turbine. This reduces fatigue and lower maintenance costs, thereby increasing availability, as well as the use of Lidar reducing costs relative to meteorological masts.

There is a problem, however, with current methods used to reconstruct the wind field from the raw Lidar measurements. Simplifications are made to ensure reasonably accurate estimation of the 10-minute average wind speed and direction [1], resulting in the following limitations which are not inherent in the Lidar sensors themselves:

- lack of reliable information about the variability of the wind conditions at higher frequencies than 10 minutes;
- poor performance when the simplifying assumptions are not valid, particularly: in complex terrain; or when the sensors are used inside wind farms, since wind conditions then vary significantly across the Lidar measurement volume due to wakes.

Machine learning offers the potential to improve significantly the accuracy and volume of wind field information reconstructed from raw Lidar data, particularly in complex conditions.

Section 2 explains the methodology. First, the physical basis for the Lidar operation and measurements, so that the inputs and desired outputs of the wind field reconstruction algorithm are made clear. Second, the machine learning method applied: Gaussian Process (“GP”) Regression. Third, the application of GPs to the wind field reconstruction algorithm.

Section 3 demonstrates the application of the developed algorithm to real wind measurements, to validate the method. Additional benefits over current methods are listed and examples given from real wind data.

Finally, section 4 draws conclusions and suggests future development priorities to realise the benefits fully and enable it to be widely used.

2. Approach

2.1. Lidar Physics

Lidar units point a laser beam (either pulsed or continuous wave) into the sky and analyse the backscatter signal to determine the radial wind speed through the Doppler effect.

Here, the Windcube Lidar unit from Leosphere is considered, which is a ground-based pulsed Lidar unit with 5 static beam directions. Four beams are oriented towards the four cardinal directions, angled by 28° from the vertical; while the fifth points vertically. Wind speeds can be measured in a 40 - 200m range, at up to 12 programmable heights (10 are generally used). All 5 beams are measured consecutively over a 4-second period.

Pulsed Lidar units have soft focus across all ranges simultaneously, since time of flight windowing can be used to detect the backscatter at a certain distance. However, since the pulse has a finite width, and because of spectral broadening, there is a measurement resolution of 25-30m in the location of the backscatter signal, which is almost constant with distance.

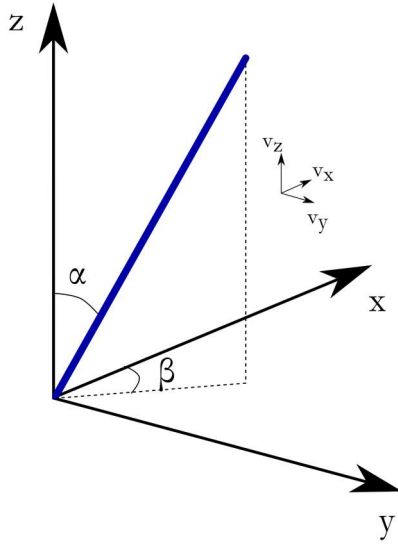


Figure 1. Spherical co-ordinate system for Lidar beam measurements.

Figure 1 shows the geometry. The component v_b of the wind velocity vector v measured by the beam (at any distance r) is the dot product of that velocity with the unit vector of the beam ($\hat{b}$):

$$\begin{aligned} v_b &= \mathbf{v}^T \cdot \hat{\mathbf{b}} \\ \therefore v_b &= -(v_x \sin \alpha \cos \beta + v_y \sin \alpha \sin \beta + v_z \cos \alpha) \end{aligned}$$

Equation 1. Radial flow speed measured by Lidar beam, due to true wind velocity.

Assumptions are therefore required to reconstruct the wind field, since only one resolved component of the wind field is measured at each point in space. This is known as the Cyclops Effect.

Currently, the standard wind field reconstruction algorithm supplied with many Lidar units assumes that all measurements within a certain time at the same height come from a stationary homogeneous process and therefore can be averaged together. For the WindCube that means averaging the North and South beams together to get v_x , and the East and West beams together to get v_y . More sophisticated methods include:

- Flow Complexity Recognition (FCR), Leosphere's patented method [2] for reproducing the wind field without homogeneity assumptions. This relies on fitting simplified Navier-Stokes equations and considering the topography of the ground around the Lidar.
- 10-minute averaged data using Taylor's "Frozen Turbulence" hypothesis [3];
- Real-time reconstruction by coupling a low-order dynamic wind model with a Kálmán filter [4];
- Reconstruction of a plane of wind velocity by using fast Navier-Stokes solver LINCOM [5];
- Use of a Wiener filter with various physics-based wind field assumptions [6];
- Use of ARMA-X modelling assuming Linear Time Invariance [7]; and
- A recent article which seems to retreat to a stationary homogeneous model [8].

The effort in [4] comes closest to using pure machine learning methods without imposing limiting assumptions on the wind field behaviour. There is however a gap for a new approach to reconstructing wind fields, minimising the number of site-specific or restrictive assumptions.

2.2. Gaussian Processes

Stochastic processes are an extension of the concept of probability distributions over values of a variable, to probability distributions over variables. In a Gaussian Process (“GP”), it is assumed that any finite number of variables described by that process are jointly Normally distributed [9]. This means that one can express a prior probability distribution over functions, which can be updated with data collected to provide predictions at any other point in the input space.

A GP is fully specified by its mean and covariance functions: $y \sim GP(\mu, \Sigma)$. A graphical overview is given in Figure 2.

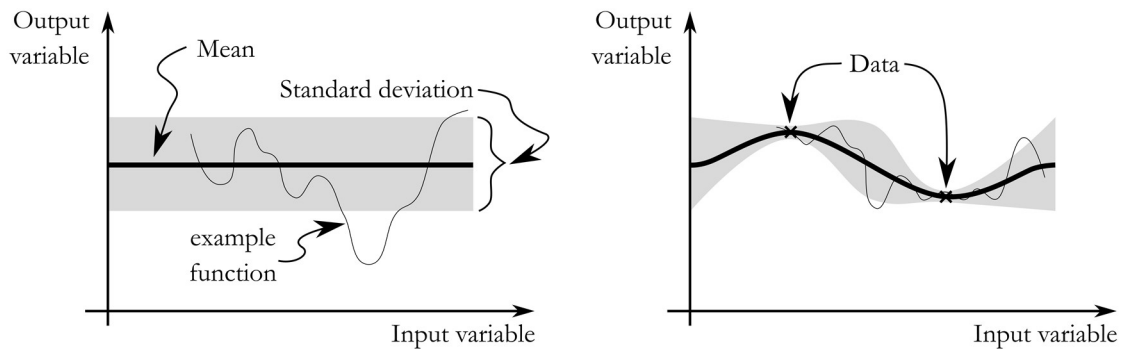


Figure 2. Basic overview of Gaussian Processes, after [9]: left, prior; right, posterior.

The covariance function k captures the relationship between the variables in the process:

$$k(x, x') = E[(f(x) - E[f(x)])(f(x') - E[f(x')])]$$

where x and x' denote values of the input variables and E is the expectation operator. Covariance “kernels” are often combined to create an appropriate model. The Matérn class of kernels provides functions which often realistically match natural processes. Each kernel is specified by a parameter ν , which is often set to half-integer values. For $\nu = 3/2$, the kernel is:

$$k_{\text{Mat}3/2}(x, x') = \sigma_f^2 \left(1 + \frac{\sqrt{3}(x - x')}{\ell} \right) e^{-\frac{\sqrt{3}(x - x')}{\ell}}$$

Once the model is specified, the hyperparameter values (e.g. ℓ) should be optimised to fit the data. The data are still required to make predictions, since this is a non-parametric model.

GP regression works particularly well when inferring information from sparse, noisy data. This class of algorithms has been applied successfully to a number of problems, including the inference of meteorological data across time and space in a country [10] and the reconstruction of topographical surfaces from Lidar scans for robotics [11] and mapping [12]. However, it has not been applied to wind field reconstruction.

2.3. Wind Field Reconstruction using Gaussian Processes

In a similar manner to that used in [11], a GP is fitted to the aspect we do not understand – the speed measured by the Lidar beam in space and time – and geometric equations are then used, along with some physical assumptions, to convert these into wind velocities. The steps in the method are:

1. Create a stationary GP (“GP1”) that can model beam velocities measured during only a small period of time (e.g. 1 minute).

2. Fit this model to a selection of periods in the data, then apply another GP (“GP2”) over the hyperparameters of those models in order to smooth them.
3. To create predictions of beam speed for a given set of time stamps, we use a set of half-overlapping GPs, with hyperparameters set from GP2 according to their time.
4. Predictions from each half-overlapping GP are combined through a form of Mixture of Experts to create the final beam velocity predictions: a “Virtual Lidar”.
5. Geometrical, physical and statistical considerations are now used to infer the wind velocity vectors from the Virtual Lidar’s beam speeds.

2.3.1. “GP1” for Short-Term Beam Speed Prediction. This GP takes measurements in space and time, in the co-ordinate system of the Lidar, and re-predicts a continuous beam speed field:

$$GP1: [\alpha_1, \alpha_2, r, t] \rightarrow v_b$$

Equation 2. Definition for GP1, mapping a static Lidar beam measurement location to beam speed.

Here, $\alpha_1, \alpha_2 \in [-\pi/2, \pi/2]$ describe the beam orientation in the well-behaved co-ordinate system $\alpha_1 = \alpha \cos \beta, \alpha_2 = \alpha \sin \beta$; $r \in [0, \infty]$ is the measurement distance from the Lidar unit; and $t \in \mathbb{R}$ is the time since a fixed reference. GP1 is defined by a zero mean function, Gaussian likelihood, Matérn 3/2 covariance kernel with Automatic Relevance Determination (ARD) [13] and box hyperpriors over the lengthscale hyperparameters for the variables α_1, α_2 and r with range [4,7].

Figure 3 shows an example output of instantaneous predictive mean and standard deviation of beam speed. As expected, a smooth variation of beam speed in space is seen, along with lower uncertainty where the measurements are taken along the beams.

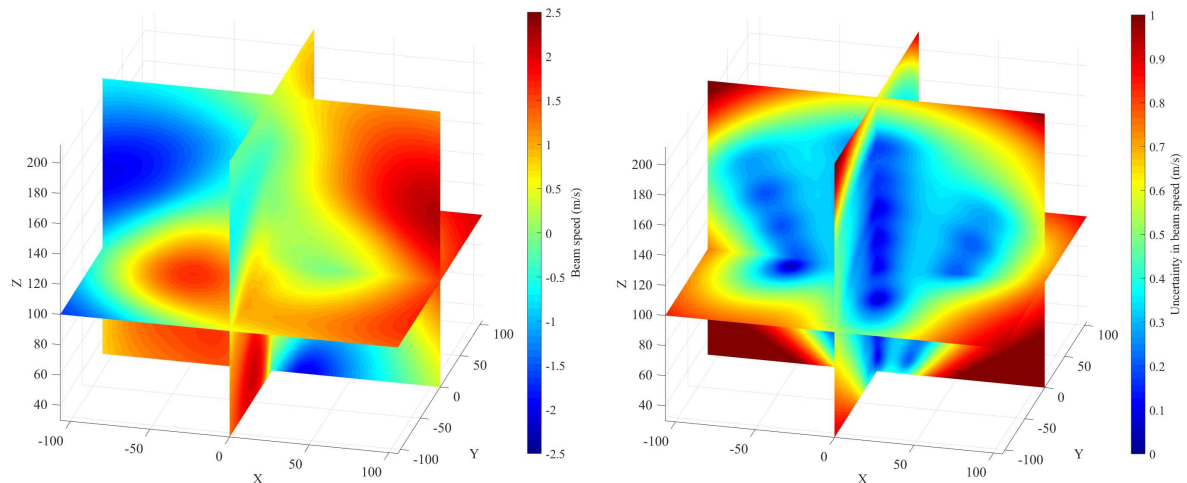


Figure 3. Plane slices across a 3D instantaneous wind field, showing mean beam speed (left) and standard deviation (right) as reconstructed by GP1.

2.3.2. “GP1” for Short-Term Beam Speed Prediction. The difficulty with GP regression in time (and in particular when fitting in real-time rather than to historical data) is that the predictions of the model are affected near the edges of the “window” by the lack of information about behaviour outside it. Wind conditions are expected to have turbulent length-scales that change in time, therefore fitting one GP for a whole day, or even 10-minute period, is not desirable.

The trade-off between amount of data available to gain a better fit and solution time is managed here by fitting one “locally-expert” GP1 to each “window” of 60 seconds of data, with a stride of 30 seconds.

As a result, each GP1 overlaps with its neighbours for 30 seconds. The individual predictions are then combined together through a weighting function, as follows:

$$v_{b*}(t) = \frac{\sum_i w_i(t) v_{b*_i}(t)}{\sum_i w_i(t)}$$

The weighting function chosen is a standard Normal distribution, centred on the middle of each GP Expert, with the time arbitrarily scaled such that the start and end are at ± 4 . This is appropriate since, as previously mentioned, the GP has more information and therefore produces more accurate results near the centre of its window. Since the predictions of each GP are effectively fully correlated, the predictive uncertainties can be calculated in the following manner:

$$\delta v_{b*}(t) = \left(\sum_i w_i(t) \right)^{-1} \sum_i w_i(t) \delta v_{b*_i}(t)$$

Setting the hyperparameters of each GP1 by independent optimisation to the local data results in unphysical behaviour. Instead, the (five) hyperparameters from local optimisation of one GP1 model every 10 minutes are obtained. A multi-output GP (“GP2”) is then fit to these values in order to smooth them, the predictions of which can then be used to set up a GP1 for any period in time covered by the data available (and predict into the future):

$$GP2: \tau \rightarrow \mathbf{h}$$

Equation 3. Definition for GP2, mapping the start time for GP1 to the hyperparameter values for GP1.

GP2 is defined by a zero mean function, Gaussian likelihood, squared exponential covariance kernel with ARD, multiple outputs specified by indicator variables, and a narrow hyperprior forcing the time length-scale to take a value of around e^{10} seconds (6 hours), creating a slowly varying function rather than one fit to the noise.

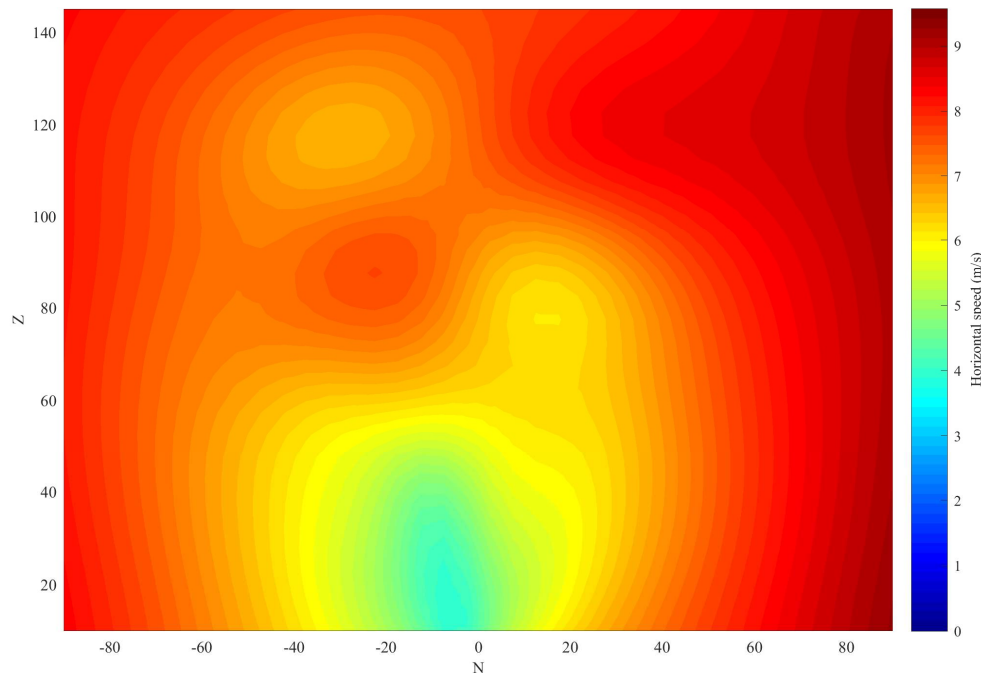


Figure 4. Vertical plane (metres East-West versus metres above ground) showing instantaneous horizontal wind speed which includes a wind turbine wake.

2.3.3. Conversion of Virtual Lidar to Wind Field. At this point, it is possible to predict (including uncertainty) the beam speed at any point in time and space, which we shall call a “Virtual Lidar”. The GPs take care of spatial and temporal variability, leaving only the issue of removing the dimensionality reduction associated with the Lidar measurement:

$$V: v_b \rightarrow \mathbf{v}$$

The simplest approach is to return to the assumptions which were earlier discarded, and obtain five points at each desired height (although now with error reduction and uncertainty estimation) and calculate one wind velocity, assuming horizontal homogeneity. This is used in the next section.

A more computationally-expensive, but valuable, approach is to obtain a number of points in any area of interest (or in an entire 3D volume) and use maximum likelihood with Equation 1 to estimate the local velocity. This provides true instantaneous 3D wind fields, meaning turbine wakes can be reconstructed, such as shown in Figure 4.

3. Results

Data from a site known to have simple wind flow conditions and where a WindCube Lidar was co-located with a met mast, as shown in Figure 5.

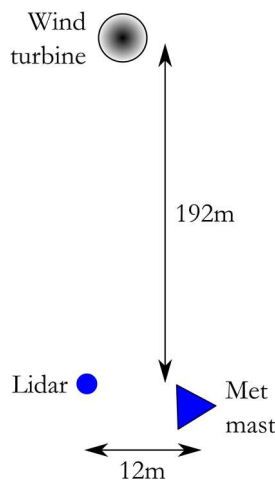


Figure 5. Site layout for validation.

The mast measurements are at 40, 60, 80, 100 and 116m, using cup anemometers and wind vanes. The mast data were only provided as 10-minute averages. The WindCube is configured to measure at exactly these heights, as well as additional heights up to 200m above ground. The time period when both Lidar and met mast data were available covers 44.6 days.

In order to resolve the Cyclops effect, beam speeds are here output from GP1 at $(x, y) = (\pm z/2, 0)$ to obtain v_x through averaging (and similarly for v_y) - very similar to the homogeneous assumptions currently made. The effect of the GP is still significant, however, since in this data set the WindCube v2 beams were not aligned with the North-South-East-West axes, but offset by 45° . The GP has therefore performed large spatial inference.

Example results are given in Figure 6 (wind speeds below 4m/s have been removed). The summary statistics for the two methods are given in Table 1 and Table 2. Bias is the mean and Scatter is the standard deviation of the residual errors.

Finally, evidence is shown in Figure 7 that GP inference during times of low data availability can result in reliable wind speed estimates.

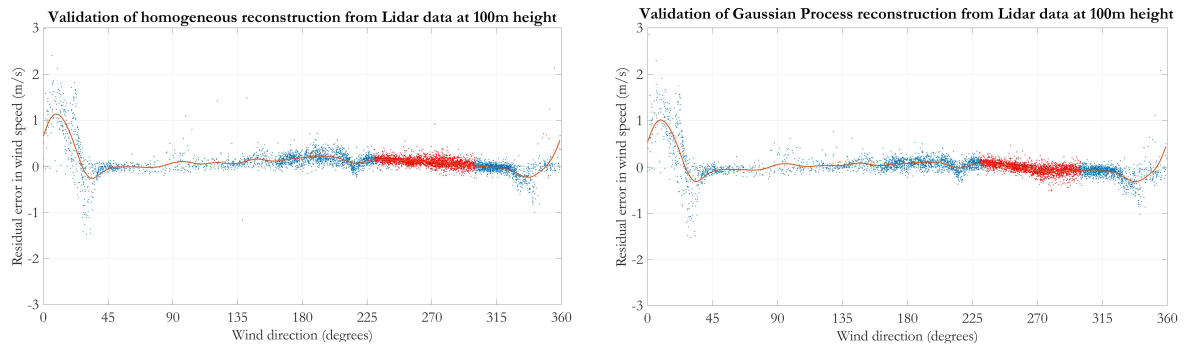


Figure 6. Diagnostic plot of residual errors versus direction at the 100m height. Left: Homogeneous (Leosphere); Right: GP. The undisturbed flow used for validation statistics are plotted in red. A smoothed average is also plotted as a line.

Table 1. Validation statistics for GP reconstruction (90% confidence interval)

	40m	60m	80m	100m	116m
Wind Speed Bias (m/s)	[-0.03, -0.02]	[-0.01, -0.00]	[-0.02, -0.01]	[-0.01, -0.01]	[-0.02, -0.01]
Wind Speed Scatter (m/s)	[0.12, 0.12]	[0.11, 0.12]	[0.11, 0.11]	[0.11, 0.12]	[0.10, 0.10]
Wind Turbulence Bias (m/s)	[-0.16, -0.16]	[-0.13, -0.12]	[-0.07, -0.06]	[-0.03, -0.02]	[-0.01, -0.01]
Wind Turbulence Scatter (m/s)	[0.12, 0.12]	[0.11, 0.11]	[0.09, 0.10]	[0.09, 0.09]	[0.09, 0.09]
Wind Direction Bias (°)	[-5.9, -5.9]	[-5.7, -5.7]	[-4.8, -4.7]	[-5.5, -5.4]	[-5.4, -5.3]

Table 2. Validation statistics for Homogeneous (Leosphere) reconstruction (90% confidence interval)

	40m	60m	80m	100m	116m
Wind Speed Bias (m/s)	[0.10, 0.10]	[0.15, 0.15]	[0.11, 0.12]	[0.11, 0.12]	[0.11, 0.11]
Wind Speed Scatter (m/s)	[0.10, 0.11]	[0.09, 0.10]	[0.08, 0.08]	[0.08, 0.09]	[0.09, 0.09]
Wind Turbulence Bias (m/s)	[-0.09, -0.09]	[-0.04, -0.04]	[-0.00, -0.00]	[0.03, 0.04]	[0.04, 0.04]
Wind Turbulence Scatter (m/s)	[0.09, 0.10]	[0.09, 0.09]	[0.08, 0.08]	[0.08, 0.08]	[0.09, 0.09]
Wind Direction Bias (°)	[-5.9, -5.8]	[-5.7, -5.7]	[-4.7, -4.7]	[-5.4, -5.4]	[-5.3, -5.3]

4. Conclusions

Lidar can be used to measure wind speeds in a volume, offering the potential for more accurate estimation of wind turbine power performance and loading. However, a Lidar can only measure the component of the wind velocity in the direction of the beam, and suffers data loss caused by clouds and rain. This makes reconstruction of the wind velocity field from raw Lidar measurements a challenge, and current methods contain assumptions which cause inaccuracies in complex terrain and in wind turbine wakes.

A novel Machine Learning method has been developed, based on Gaussian Process regression, to remove these assumptions. This approach is naturally robust to overfitting, predicts uncertainty derived from data density and machine error, and – subject to further validation – is applicable to any site without tuning. Further, the method can be applied to most of the monostatic Lidar designs, with some adjustments for circular scanning patterns and in the future for floating Lidar.

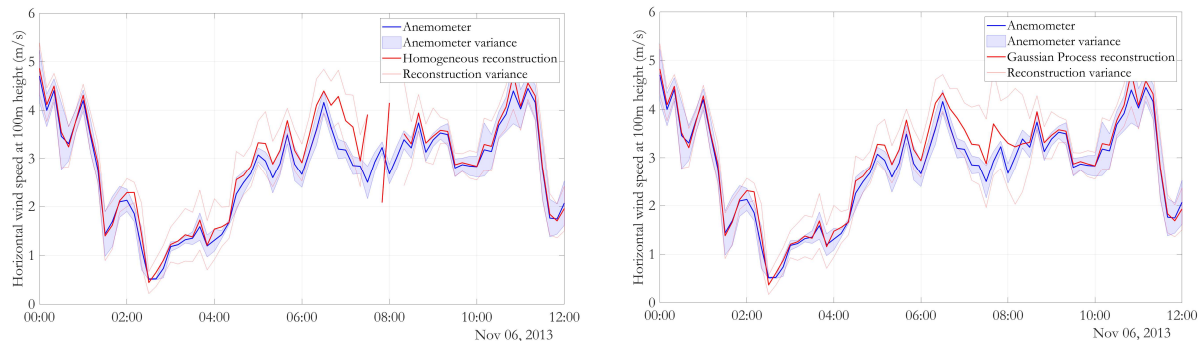


Figure 7. Comparison of 10-minute averaged wind speed from the met mast (blue) with reconstruction from Lidar (red) during a period of low data availability. Left: Homogeneous (Leosphere); Right: GP.

Validation at an onshore site with simple wind conditions over 44.6 days, against a met mast with cup anemometers and wind vanes at 5 heights covering 40–116m, gives the following results:

- The new method achieves an average wind speed bias of -0.01m/s and a scatter of 0.11m/s , meeting the threshold for commercial use.
- The new method infers data during Lidar measurement gaps, offering the potential for 100% data availability.
- Errors at one measurement height are automatically removed.
- Instantaneous 2D wake measurements in a plane can be made, which when validated and further developed will offer a breakthrough in understanding of wind farm interactions.

Despite these promising results, some priorities have been identified for future development:

- Conduct validation against 1-second sonic anemometer data.
- Further improve methods used for creating the Virtual Lidar.
- Adapt the model for floating Lidar by correcting for the rotation, and potentially adding additional terms for the translation, in GP1.
- Further improve methods used for reconstructing wind velocity from the Virtual Lidar.

References

- [1] Bossanyi E, Kumar A and Hugues-Salas O 2014 Wind turbine control applications of turbine-mounted lidar *J. Phys. Conf.* **555** 012011
- [2] Boquet M, Nibart M and Albergel A 2014 Method and device for determining the movements of a fluid from remote measurements of radial velocities *Europe Patent* EP2661636
- [3] Raach S, Schlipf D, Haizmann F and Cheng P W 2014 Three dimensional dynamic model based wind field reconstruction from lidar data *J. Phys. Conf.* **524** 012005
- [4] Towers P and Jones B 2014 Real-time wind field reconstruction from lidar measurements using a dynamic wind model and state estimation *Wind Energy* **19** 133–50
- [5] Mikkelsen T, Herges T, Astrup P, Sjöholm M and Naughton B 2017 Wind field re-construction of 3d wake measurements from a turbine-installed scanning lidar *abstract from Int. Conf. on Future Technologies for Wind Energy* (Boulder: DTU)
- [6] Guillemin F, Domenico D, Nguyen N, Sabiron G, Boquet M, Girard N and Coupiac O 2016

- Nacelle lidar online wind field reconstruction applied to feedforward pitch control *J. Phys. Conf.* **753** 052019
- [7] Givanani A, Bierbooms W and van Bussel G 2016 Estimation of rotor effective wind speeds using autoregressive models on lidar *J. Phys. Conf.* **753** 072018
 - [8] Borraccino A, Schlipf B, Haizmann F and Wagner R 2017 Wind field reconstruction from nacelle-mounted lidar short-range measurements *Wind Energy Science* **2** 269–83
 - [9] Rasmussen C E and Williams C K 2006 *Gaussian Processes for Machine Learning* (The MIT Press)
 - [10] Tastu J, Madsen H and Pinson P 2013 Short-term wind power forecasting: probabilistic and space-time aspects *DTU PHD-2013* **306**
 - [11] Smith M, Posner I and Newman P 2010 Generating implicit surfaces from lidar data in *Towards Autonomous Robotic Systems* (Plymouth, UK)
 - [12] Liu Y, Monteiro S T and Saber E 2015 An approach for combining airborne lidar and high-resolution aerial color imagery using gaussian processes in *Proc. Image and Signal Processing for Remote Sensing XXI* (Toulouse: SPIE) vol. 9643 ed. L Bruzzone
 - [13] Neal R M 1995 Bayesian Learning for Neural Networks *Lecture Notes in Statistics* vol. 118