

Interactive comment on “Parameterization of Wind Evolution using Lidar” by Yiyin Chen et al.

Yiyin Chen et al.

chen@ifb.uni-stuttgart.de

Received and published: 24 June 2020

First of all, we would like to thank all the reviewers for their time taken to read our manuscript and their constructive comments. We have considered all the comments in detail and revised our paper accordingly. We believe that these comments have helped us to further improve the quality of our paper.

Please find below our responses to reviewer comments. The reviewer comments are repeated in black text, and our responses are provided in blue text.

C1

Response to comments of Felix Kelberlau

General comments

Chen et al. develop a method to predict the coherence of horizontal wind velocity fluctuations for mostly longitudinal separations. Their predictions are based on first to fourth order wind speed statistics that can be calculated from either nacelle-mounted lidar or mast-based in-situ anemometry. They use data from two measurement campaigns to test their approach and find good results that are especially relevant for lidar-assisted wind turbine control. The work lies therefore well in the scope of WES and is of broad international interest. The paper builds up on an existing wind evolution model and presents a novel approach to parameterise its two coefficients by means of machine learning. The manuscript explains the study thoroughly and reproducibly, presents all relevant results and discusses them critically.

[We would like to thank the referee for the interest in this research.](#)

Section 2.6 "Gaussian Process Regression" lies outside my field of expertise and I can therefore not evaluate if the chosen model is suitable for the task of parametrization the wind evolution model.

[Gaussian process regression is a powerful modeling tool. One of the objectives of this paper is to explore if this method can be applied to wind evolution modeling and the results have demonstrated its potential.](#)

The manuscript is overall understandable but would benefit greatly from being proof-read by a native English speaker or similarly qualified person before publication. I recommend reconsideration for publication after major revisions.

C2

Thank you for your suggestion. The revised version will be proofread before submission.

Specific comments

l. 2: I assume you mean "the mean flow" (also l.13 and all other occurrences).

Yes. This has now been corrected throughout the paper.

l. 50: The introduction would benefit from references to research that support Taylor's frozen turbulence for very large turbulent structures but limit its applicability for long separation distances or small scale turbulence such as Willis and Deardorff (1976), Schlipf et al. (2010), and Kelberlau and Mann (2019).

Thank you for your suggestion. We will mention the relevant research in the introduction.

l. 59: "the vertical intercept" It would be better to describe the second parameter without referring to the coherence-frequency plot that is not yet introduced here.

Thank you for your suggestion. This sentence has now been rephrased.

l. 61: "Mann spectral velocity tensor" Mann (1994) should be cited here.

Thank you for your suggestion. The citation has been added to the text here.

l. 68: "If any data... is also available..." Please mention which data is available or would be of interest.

We noted this sentence is not well formulated and have now rephrased it. And the data used in this research is introduced in Sect.3.

l. 94: Please introduce this travel time as a function of the mean wind speed

C3

here.

In this context, the travel time Δt is in a general sense, not specifically the travel time approximated by d/U , which is defined as Δt_{Taylor} in l.130.

l. 97: Please explain why "it is not possible to predict every point of the coherence curve". For my understanding, the coherence curve is visible on a plot like in Fig.1. Do you refer to not having not enough data to smoothen the curve or not having data for all separation distances?

In this sentence, the *coherence curve* means the *estimated coherence curve*, which is the "reality" we aim to approach. And the smooth coherence curve is acquired from a *wind evolution model*. We think it is important to clearly distinguish the different meanings of *wind evolution* and *wind evolution model*. Wind evolution is a physical phenomenon (the "reality") and wind evolution model is a model used to approximate it (There are different wind evolution models). It is not possible to predict wind evolution but to predict the parameters of a wind evolution model and use this model to approximate it.

In fact, we'd like to explain our prediction concept at an abstract level in Sect. 2.2. We think that the key to using machine learning to build predictive models is to find suitable *predictors* and *targets* — This is the process of abstracting and condensing information. Essentially, using a wind evolution model is to condense the information in the estimated coherence into several model parameters, which are predictable. We will improve this section to make the logic more understandable.

l. 100: Do you mean "...according to measured wind velocity time series by a parameterisation model"?

"Wind field conditions" here refer to all variables related to a wind field, not limited to statistics of measured wind velocity time series. But we noted that this word might be not precise enough. We will explicitly write down the types of relevant variables instead.

C4

Fig. 1 and Fig. 2: In general, it is good to visualize the workflow like done here. But both figures show overlapping information and I recommend to merge them into one figure. The numbering used in Fig.2 with an explanation in the text and caption(!) is more informative than the keywords currently used in Fig. 1. A figure and its caption should be self explanatory whenever possible. Please try to improve the text l. 98-106 for better understanding.

Thank you for your suggestion. We will merge Fig. 1 and Fig. 2 into one figure and modify Sect.2.2 accordingly.

l. 120: Please describe which frequency you are referring to. Probably the frequency of the horizontal wind velocity fluctuations.

Yes, it is the frequency of horizontal wind velocity fluctuations. We've made it clear in the text.

Maybe also introduce the wavenumber k here that is used as a measure of eddy size in many other publications.

Thank you for your suggestion. We agree that wavenumber is also a common measure in spectral analysis of turbulence. However, it is not used in this paper because our study is based on dimensionless frequency. But we think it makes sense to mention the relationship between wavenumber and dimensionless frequency.

Eq.(5): It is not clear where (5) comes from. If you do not want to include the complete deduction, I suggest to give a reference that shows it and uses the same form of the equation. In Simley and Pao (2015), a and b are defined a bit differently, I think.

We will include the complete deduction and clarify the reason for adapting the equation of Simley and Pao (2015) in the revised manuscript.

l. 147: You should include the weighting here: e.g. "...but the weighted average of the wind speeds within the measurement volume"

C5

Thank you for your suggestion. The corresponding text has now been modified.

l. 148: This is a bit ambiguous because spatial averaging does also refer to combining data from different measurement volumes in different lidar beam directions. Better write: "so-called line-of-sight averaging effect of lidar". (also l. 156)

Thank you for your suggestion. The corresponding text has now been modified.

l. 151: Please refer to more fundamental work (Nyquist-Shannon sampling theorem).

Thank you for your suggestion. The following reference has been cited: C. E. Shannon, "Communication in the Presence of Noise," in Proceedings of the IRE, vol. 37, no. 1, pp. 10-21, Jan. 1949, doi: 10.1109/JRPROC.1949.232969.

l. 152: You should mention the sampling rate of the lidars here (not only in the table) and compare it with the frequency of the eddies that you want to detect.

According to the paper structure, Section 2 introduces the theoretical framework of this study and gives related discussions in a general way. Measurement-related content is first introduced in Section 3.

l. 158: The line-of-sight weighting of a pulsed lidar is usually approximated by a triangular function as in e.g. (Sathe and Mann (2012)) which is a sinc²function in the frequency domain.

Thank you for your suggestion. We agree that weighting functions for the volume averaging effect of pulsed lidars could have different functional forms. We've added the triangular function as an example of weighting functions.

Indeed, the functional form of the weighting function mainly depends on the shape of emitted pulses and the sampling of backscattered pulses. For example, a triangular function is used for the case where the pulse shape is assumed to be ideal rectangular (Sathe and Mann (2012)). For Leosphere pulsed lidar systems, a Gaussian weighting

C6

function is usually used, see e.g. the following reference:
 Carious, J.-P.: Pulsed Lidars, in: Remote Sensing for Wind Energy, DTU Wind Energy-E-Report-0029(EN), chap. 5, pp. 104–121, 2013.

I. 184: It should be considered that $w(x)$ is approximately 0 for fluctuations that occur with a wavelength of twice the length of the illuminated section of the lidar beam (or length of the range gate). In this case the measurement signal would be determined by noise only. I suggest to estimate a range of critical frequencies based on the length of the range gates. This range of critical frequencies should be considered in the further analysis, if it is relevant for the results.

Thank you for your suggestion. That is a good point to check.

According to Schlipf (2015), the critical wavenumbers are $2\pi/W_L$ (W_L is the full width at half maximum) and its harmonics. The relationship between wavenumber and dimensionless frequency is $kd/2\pi$. Thus, the first critical dimensionless frequency is d/W_L . For example, consider $W_L = 30$ m (for Leosphere systems) and $d = 27.25$ m (the smallest separation of LidarComplex, which is the most critical case), $d/W_L \approx 0.91$, which is located in the filtered part (the grey area). Therefore, the critical dimensionless frequency is not relevant for the results. This discussion will be briefly mentioned in the related part.

Schlipf, D.: Lidar-assisted control concepts for wind turbines, Dissertation, 2015.

Your derivation assumes furthermore that the weighting function is identical for all range gates. This is only true if the laser beam is well collimated. Is this the case for the lidar devices used in this study?

We agree that assuming identical weighting functions for all range gates is a simplification. As mentioned in the paper, the derivation is based on ideal assumptions.

I. 199: You could mention that a lidar with additional beams would help here and could also be used to avoid yaw-misalignment.

Thank you for your suggestion. This has now been added to the corresponding text.

C7

I. 208: What is the expected order of magnitude for the misalignment angle?

First of all, we must emphasize that we don't have the data of turbine misalignment. What we have is the yaw position of the turbine and the wind direction measured on a met mast located 295 m away from the turbine. When calculating the misalignment angle α , the mean wind direction at the turbine is approximated with the mean wind direction measured on the met mast (please note the possible uncertainty). α is approximately normal distributed, with $\sigma \approx 5^\circ$.

How much "decorrelation" do you expect from a turbulence model (e.g. Mann (1994)) due to the resulting lateral separation? Can you quantify the order of magnitude of the resulting error approximately?

First of all, we want to emphasize that we did not ignore the influence of the misalignment angle on the horizontal coherence, but defined it as a predictor.

The resulting lateral separation depends on the separation between two range gates. Here, we make a simple comparison based on the coherence model of Kaimal spectrum:

$$\gamma(r, f) = \exp \left[-12 \cdot \sqrt{\left(\frac{f \cdot r}{V_{\text{hub}}}\right)^2 + \left(0.12 \frac{r}{L_c}\right)^2} \right]. \quad (1)$$

We find one data block where $\alpha = 0^\circ$ and compare the longitudinal coherence estimated from this data block with the theoretical lateral coherence calculated according to the above equation, assuming $\alpha = 5^\circ$ (see the attached figure). Two measurement separations are included, $d_1 = 27.25$ m and $d_2 = 81.75$ m. The mean wind speed $V_{\text{hub}} = 11.7$ m/s.

I. 213: Please always write which variable you are referring to when you mention standard deviation sigma.

Thank you for your suggestion. This has been improved.

I. 236: It would be good to introduce the variable alpha already in 2.4 (I. 199

C8

and Fig.4) if you refer to it here.

Thank you for your suggestion. The introduction of alpha has now been moved to Sect.2.4.

I. 263-266: This sentence is very long and difficult to understand.

Thank you for your suggestion. We will try to rephrase it to make it more comprehensible.

I. 332-343: Please reassess which information should be given here: I miss: the measurement height of the lidar, length of each range gate, measurement distances...Some of these values are given in Table 2 but should also appear here.

Thank you for your suggestion. We've now added these information to the corresponding text.

The information about the coordinate system will not be used again later in the text and do not need to be given at all then.

The wind coordinate system is defined for the processing of sonic data. Sonic data from LidarComplex is used in the analysis (see Table 5).

I. 342 and 350: Main wind direction refers usually to the direction from where the wind blows most frequently. Better write mean wind direction.

Thank you for your suggestion. This has been corrected throughout the paper.

I. 389: I suggest a similar filtering against the line-of-sight averaging. See comment for I.184. Probably it is not worth it to re-run the computations. But check in the coherence plots, if the frequency range is relevant and if you see a random increase in coherence in it.

Please find the response to the comment for I.184.

I. 397: Why do you not also filter the lowest percentile?

C9

Because the value distributions of the parameters all have a long right tail. We've added this explanation to the text.

Fig. 6: Subfigure (e) would benefit from a zoom into just some few minutes of data with thin plot lines to show all velocities clearly and not on top of each other.

Thank you for your suggestion. We'd like to show the whole 30-min data block because the coherence is estimated from it. We will try to use thinner lines to make the curves more visible.

I. 445: That makes the difference caused by the reduced sampling rate in Fig. 7 (b) even more interesting. Earlier you write that "As long as the sampling rate of lidar is sufficiently high to acquire a complete coherence curve, it will not have a noticeable effect on the study of the coherence." but here you find that the influence on parameter b is big (logarithmic y-axis). What could be the reason? $f_s = 1/3$ Hz means, at least at high wind speeds, that it is difficult to measure exactly after the eddy travel time passed. That means you might miss the best moment to take your measurement. Could this maybe have to do with the results?

Thank you for sharing your opinion about the possible influence of a low sampling rate. However, in this study, we found the difference caused by the reduced sampling rate too small to be observed in a plot with a normal y-axis. Both curves completely overlap. That is exactly the reason for using a logarithmic y-axis instead. Please note that a logarithmic axis will enlarge the difference between values lower than one.

I. 462: This is a very interesting finding! Better write that the coherence is dependent on the separation distance but independent of the measurement distance, i.e., the position in the induction zone.

Thank you for your suggestion. The sentence has now been improved.

I. 471 ff.: You point out that "The decay parameter a shows a decreasing trend

C10

with increasing measuring separation." After reading the explanation given in lines 475–480 several times I still do not fully understand why longer travel time (or separation distance) leads to a less decay.

Because the modeled coherence curve is determined by a and b together. b increases with separation between the two observed points, which means the coherence at zero frequency decreases with separation. In other words, "the starting point" of the coherence curve is lower, and thus the room for coherence decay becomes smaller.

And if I could follow the explanation, it would not explain why the value increases again for very long separations with ParkCast data. Maybe you can explain this better?

In our opinion, a decreases with separation, and its decreasing trend gradually stops at a separation of around 300 m (see Fig.9 (c)). We are not sure why you think "the value increases again for very long separations with ParkCast data". If it is because of the upper whiskers, we think this indicates the value is more scattered for long separations. When we observe the trend, we mainly focus on the value range of 25th to 75th percentiles (the box) and the median value.

I. 489: You should explain the results shown in Fig. 11 at least briefly to justify your predictor selection shown in Table 5. Please mention the $\log(\sigma_m)$ thresholds.

Thank you for your suggestion. We will explain the predictor selection in more detail.

Fig. 12: The scatter plots are not as informative as they could be. Please decrease the marker size and maybe add transparency to make it possible to see the density of the data points. Also a regression line would help to quantify the relation between x and y .

Thank you for your suggestion. We will improve Fig.12 according to your suggestion.

Table 5: Please provide a better caption. What does for example bold font mean?

Thank you for your suggestion. The caption has been improved.

C11

I. 564: The prediction accuracy is very good for this one example case but from Fig.12 we know that the scatter throughout the whole dataset is quite high. Is this particular example (12.12.2013, 12:00–12:30) representative in any way?

We chose this data block based on two principles: 1) data integrity and 2) representative wind statistics. In this example, the lidar measured mean wind speed is 7.3–7.7 m/s and the lidar measured turbulence intensity is 0.10–0.12, for different range gates. These values appeared frequently in the selected period according to Fig.A1. Hence, we think this example is representative for the data involved in this study. And we decided to choose a data block as an example because we thought neither the coherence itself nor its prediction are intuitive (esp. a and b are predicted respectively but must be combined to give the modeled coherence). We noted a mistake here: the date should be 07 Dec. 2013. Figure 6, 8, and 13 are all plotted with the same data block. The date has now been corrected.

Maybe it is more interesting to show a plot for a case where the deviation between modelled and fitted curves equals the RMSE?

Thank you for your suggestion. Maybe we could try to plot two additional curves of the modeled curve \pm RMSE in the example plot to indicate the range of RMSE.

I. 599: The theory about atmospheric stability in section 2.5 can be removed then.

We did analyze the influence of atmospheric stability. However, the stability happens to be mostly neutral in the period we chose, and thus we could not get any clear conclusions from the data. We will add an explanation about this issue in the text.

I. 625: The computation time was not mentioned before. Rather don't mention it only in the conclusions.

We have found out in this research that computational time could be a matter of concern when applying machine learning methods. Thus, we'd like to suggest it as one of the topics worth studying in the future in the outlook.

C12

Fig. A2: Please add a grid and either add x-ticks or if that is not possible out of confidentiality concerns, remove [Hz] and [m/s²] from the labels.

Thank you for your suggestion. The grids and x-ticks in Fig.A2 were intentionally removed out of confidentiality concerns. We will remove the units as well based on your suggestion.

Technical corrections

l. 16: "taking values", "1..."

l. 45: "lidar is a remote sensing technology"

l. 67: Remove "Some"

l. 81: Remove "again"

Fig. 1: The text in the plot is too small. Maybe simply enlarge the whole plot a bit.

Fig. 2: The caption is a stub.

l. 109: "...with only a few simple parameters." or better: "... with as few parameters as possible."

l. 114: "a linear function or a more complicated term."

l. 127: Please check all occurrences of "the both" and use either "the two" or only "both" instead.

l. 131: "dimensionless frequency" should be written in roman script.

l. 141: "projected onto"

l. 150: Please check all occurrences of "starring" and change them to "staring".

l. 151: Please capitalize "Oppenheim".

l. 155: "...laser beam pointing into a fixed direction."

l. 200: "Since..." the sentence is not correct. Please rewrite.

l. 234: "So as the travel time Dt" is not a sentence. Please rewrite.

Table 1: Unusual table style. Consider three columns including a header for each column instead of lines separating the rows.

C13

l. 351: "sampling rateS because of THEIR"

Table 2: The Matlab like interval syntax is maybe not the best way to present the measurement distances. Better write, e.g., 30,90,...,990, Dec is abbreviated, June is not.

l. 375: What does C stand for? Probably a formatting error here?

l. 379: and 381: No new paragraphs here.

l. 399: "referred to throughout"

l. 419: Remove both occurrences of "as"

l. 459: The idea for colours and markers is good but most of it is not visible in the plot. Try slightly thinner lines and different marker sizes for different colours (e.g. blue circle tiny, red circle small, yellow circle medium...).

Figure 9: The caption does not explain the difference between a) and c) and between b) and d).

l. 579: "And the more noisy..."

l. 586: "nacelle-based"

l. 611: "error is" or "errors are"

We'd like to thank the reviewer again for these suggestions for technical corrections. We have considered these comments in detail and made corresponding corrections and improvements.

References

Willis and Deardorff (1976): 10.1002/qj.49710243411

Schlipf et al. (2010): 10.18419/opus-3915

Kelberlau and Mann (2019): 10.5194/amt-12-1871-2019

Mann (1994): 10.1017/S0022112094001886

Sathe and Mann (2012): 10.1029/2011JD016786

C14

Response to comments of Anonymous Referee 2

General comments

This manuscript presents a statistical model of longitudinal coherence describing the evolution of turbulent structures in the wind as they travel downstream. The topic is very relevant for lidar-assisted control applications (and other wind preview-based control applications), where a good understanding of the correlation between the wind at the measurement point and the turbine is needed. There has been previous work in the literature focusing on developing wind evolution coherence models with parameters describing atmospheric conditions as inputs. However, the existing models don't necessarily fit observed data well for all atmospheric conditions. This manuscript includes many additional atmospheric parameters as predictors to estimate the coherence and also applies a machine learning approach to model wind evolution. The advantage of the machine learning approach is that the set of parameters used to predict wind evolution can be adapted to the measurements available at a given location. The manuscript describes novel and relevant research, and overall is well written.

We would like to thank the referee for the interest in this research.

Despite the significance of the research, there are several areas that I believe should be addressed. First of all, it would be useful to understand how the accuracy of the developed model compares to existing wind evolution models (e.g., Kristensen, 1979; Simley and Pao, 2015; possibly Davoust and von Terzi, 2016). The manuscript claims that the developed model is sufficiently accurate to model wind evolution, but if possible, it would be interesting to know how much it improves over these simpler models.

Thank you for your suggestion. Indeed, we have also considered to compare the results of the GPR models and that of some existing wind evolution models. However, our

C15

concern is if this kind of comparison would make sense given the different conditions in our study in comparison to the others. Here, we would like to take the results of Simley and Pao (2015) as example, because the wind evolution model used in this research is adapted from that one. Firstly, the curve fitting is done differently. In Simley and Pao's work, the objective function for fitting is the sum of the squared errors weighted by the corresponding power spectrum (See the equation (5) and the corresponding explanation in Simley and Pao (2015)). However, in our work, no weighting function is applied in the fitting. Therefore, the fitted coherence curves will be slightly different in both cases even for same data, and thus the corresponding model parameters will be slight different as well. Secondly, in Simley and Pao's work, the input variables used to determine the model parameters are supposed to be acquired from ideal point measurements because the model is developed from LES data. However, it is not possible for us to acquire equivalent input variables from the on-site measurements. Despite these difficulties, we will look into the possibility of making a comparison again.

Second, the manuscript is very well organized and easy to follow! But the English usage could be improved throughout the manuscript. For example, there are several sentence fragments, the word "the" is used in many places where it is not needed, and some of the language seems too casual (e.g., pg. 10, ln. 255: "Think of making a regression model from some data.").

Thank you for your suggestion. The revised version will be proofread before submission.

My biggest concern with the manuscript is that the analysis assumes that the spatial averaging effect of the lidar can be ignored (discussed on pgs. 7 and 8). The authors correctly show that the lidar weighting function does not affect the measured coherence as long as it is assumed that wind evolution can be ignored within the probe volume (Taylor's hypothesis is applied). But this over-simplifies the problem. For example, the

C16

authors are estimating the wind evolution between the two adjacent range gates, with range gate spacing as low as 27.5 m. But pulsed lidars typically have a Full Width at Half Maximum width of ≈ 30 m. Therefore, it seems problematic to assume Taylor's hypothesis within the 30 m probe volume, but assume wind evolution between the two range gates separated by a similar distance. From my own analysis of the impact of spatial averaging on the measured coherence, when the wind evolution model is applied within the probe volume as well as between range gates, the presence of the weighting function significantly impacts the measured coherence. This has the effect of increasing the low frequency coherence but causing the high frequency coherence to decay much faster. Therefore, it seems likely that ignoring spatial averaging altogether in this work leads to incorrectly fitting the coherence model.

The authors should include some analysis comparing the modeled coherence with and without wind evolution within the probe volume, using the results to either justify their approach or to show that Taylor's hypothesis cannot be ignored. A better approach would be to include the impact of spatial averaging and find the a and b parameters that best fit the measured coherence when the wind evolution model is combined with the spatial averaging model. In principle, this approach is similar to the method developed by Schlipf et al., 2015 (*Meteorologische Zeitschrift*), but much simpler since only a staring lidar mode is used.

Thank you for pointing out the interesting question about wind evolution within the probe volume. We'd like to explain our consideration about this issue as following: In principle, wind evolution depends on the evolution time of turbulence (see equation (2) in the paper). Theoretically, Taylor's hypothesis is valid as the evolution time approaches zero. Although the probe volume seems to have a similar length as the distance between two adjacent range gates, the corresponding evolution time of both cases is totally different. The typical length of a laser pulse is in the order of magnitude of 10^{-7} s (e.g. 150-400 ns depending of devices). This temporal length corresponds to a spatial length in the order of magnitude of 10^1 m considering the light speed. This

C17

means it is not possible for lidar to distinguish the signals backscattered from the locations within this spatial range. This is the reason for a lidar having a probe volume. In this case, the evolution time of turbulence is in the order of magnitude of 10^{-7} s. But for the distance between two range gates, the evolution time corresponds to the travel time of the mean flow between two range gates, which is in the order of magnitude of $10^0 - 10^1$ s depending on wind speed and the distance between both range gates. Therefore, we think assuming Taylor's hypothesis within the probe volume is reasonable. We will add this explanation in the corresponding part of the paper to avoid misunderstanding.

Nevertheless, we agree that ignoring the spatial averaging effect of lidar is based on ideal assumptions (esp. the laser beam aligns with the mean direction, which is not always the case in practice) and is a kind of simplification. And in our paper, we also suggest the method developed by Schlipf et al., 2015 (*Meteorologische Zeitschrift*) for cases where the misalignment angle between the mean wind direction and the laser beam can be determined accurately (because this method requires the misalignment angle). In fact, our deduction follows the same approach but assumes no misalignment angle. The reason for that is, as discussed in the paper, determination of the misalignment angle is not always possible. For example, in our case, we can only use the wind direction measured on a met mast located at about 300 m away from the wind turbine to approximate the wind direction at the wind turbine. This approximation contains uncertainties. And sometimes even if it is possible to acquire the misalignment angle at turbines, the requirement for accuracy is very high because this variable is included in the most basic step — fitting the estimated coherence to the wind evolution model. Based on these considerations, we decided not to include the angle in the fitting but use it as a predictor, which makes this variable more standalone and prevents its errors from affecting "everything". And Gaussian process regression inherently assumes imperfect training data (containing noisy terms). Thus, it is better to keep uncertainties in predictors.

C18

Moreover, in this research, our goal is to explore the potential of applying GPR in prediction of wind evolution. We wanted to examine it with different data availability. Thus, we could not include the misalignment angle in the fitting process, assuming this variable is always available. And we wanted to use a simple wind evolution model as a baseline case to demonstrate the prediction concept. As mentioned at the end of the paper, one could choose whichever wind evolution model suitable for own application scenario and obtain the corresponding parameterization model by following the methodology suggested in this work.

Specific comments

-Pg. 2, In. 58: "adapted the Pielke and Panofsky's model by introducing a new parameter..." More accurately, the paper by Simley and Pao (2015) took the form of the coherence model for transverse and vertical separations suggested by the following paper, and adapted it to longitudinal coherence:

R. Thresher, W. Holley, C. Smith, N. Jafarey, and S.-R. Lin, "Modeling the response of wind turbines to atmospheric turbulence," Department of Mechanical Engineering, Oregon State University, RL0/2227-81/2, Corvallis, OR, Tech. Rep., Aug. 1981.

Thank you for your suggestion. This reference has now been cited to the corresponding text.

-Section 1: Introduction: Another very relevant paper should be discussed in the literature review section. The following paper discusses fitting lidar-measured coherence to the longitudinal coherence structure suggested in Simley and Pao (2015):

Analysis of wind coherence in the longitudinal direction using turbine mounted lidar S.Davoust and D. von Terzi 2016 J. Phys.: Conf. Ser. 753 072005

Thank you for your suggestion. This paper is very relevant to our work. A short discussion about it has now been added to the literature review.

C19

-Eq. 9: The last index "j" should be changed to "i".

Thank you for pointing out this mistake. It has now been corrected.

-Pg. 8, In. 188: This paragraph and Fig. 4 are hard to follow. I would suggest labeling the angles the text refers to in the figure, and also provide some equations to support what you are trying to explain.

Thank you for your suggestion. We will improve the figure and the corresponding text.

-Pg. 8, In. 196 - pg. 9, In. 208: In this discussion, it is a little hard to tell if yaw misalignment is required by the coherence estimation method, or if it is optional. This becomes obvious later, but I think here it would be good to explain that the final model allows different combinations of predictors (including yaw misalignment) depending on availability.

Thank you for your suggestion. We will improve this part and mention that the misalignment angle could be used as a predictor if it is available.

-Section 2.5: Can you compare the predictors you are using to the predictors used in previous longitudinal coherence models in the literature (e.g., Kristensen, 1979; Simley and Pao, 2015)? It would be insightful to understand which new parameters are included in this study.

Thank you for your suggestion. We will add a simple comparison of predictors to this Section.

-Pg. 9, In. 213: It would be good to define turbulence intensity here.

Thank you for your suggestion. The definition of turbulence intensity has been now added to this part.

-Pg. 10, In. 233: "thus how likely or to what extent the local terrain changes" makes it sound like the terrain variations are the primary reason the coherence would

C20

depend on "d". But even if the terrain stays the same, I would still think there could be a dependence on "d".

Thank you for pointing out this issue. Actually, we don't want to imply the terrain variations are the primary reason. But we have to admit the expression in that paragraph is not good enough and thus causes this misunderstanding. We will modify this paragraph to make it clearer.

-Pg. 10, In. 234: "For prediction, it is not possible to obtain $\Delta t_{\max\text{corr}}$." Why can't it be determined? It can be calculated just like all the other predictors, right?

Yes, you're right. Thank you for pointing out this mistake. We will include $\Delta t_{\max\text{corr}}$ as a potential predictor and add a short discussion about selecting $\Delta t_{\max\text{corr}}$ or Δt_{Taylor} in Sect.5.2.

Also, on pg. 6, In. 130, you say that the d/U approximation is not used in this study and $\Delta t_{\max\text{corr}}$ is used. Which of these statements is right?

Thank you for pointing it out. This is a mistake in writing. The sentence in l.130 should be "... this approximation is not applied in estimation of coherence." When estimating the coherence, the velocity time series measured at the downstream is shifted by $\Delta t_{\max\text{corr}}$ (l.377), while Δt_{Taylor} could be used as a predictor considering it is easy to calculate.

-Pg. 10, In. 251: (Chen, 2019). Can you describe how this current manuscript compares to the earlier work? Better yet would be to discuss this in the introduction.

Thank you for your suggestion. A brief introduction of the preliminary work has been added to the introduction.

-Pg. 11: Section: "Hyperparameters of GPR": In general, this section would be more clear if the specific variables discussed were connected to the wind evolution application.

C21

This part aims to introduce the hyperparameters of GPR in a general sense so that readers could more or less understand the functions of these hyperparameters. In general, the logic of machine learning is to find statistical relationships among data (if we say it in a simple way). It is not possible to associate its algorithm to specific physical quantities except predictors (inputs of models) and target variables (outputs of models), which are introduced in other sections. Thus, we think it would be better to keep the explanation here abstract.

-Pg. 11, In. 271: "where x is the input vector of different parameters" Can you provide an example of what these input parameters are in your application?

The input vector x is a set of predictors for a single observation, which has the dimension of $D \times 1$. We've added the dimension of predictors D to the text.

-Pg. 12, In. 278: "where X is the aggregation of all input vectors." Can you explain in more detail? What are the dimensions of X ? of parameters x of observations?

X has the dimension of $D \times n$. D is the number of predictors and n is the number of observations. We've added the number of observations n to the text.

-Eq. 19: I did not see these basis functions or the coefficients beta discussed any more in the manuscript. Can you describe how you chose the basis functions and how the coefficients were estimated? And how do these values affect the final estimate in Eq.22?

There are four types of basis functions provided in MATLAB: empty (assuming no basis function), constant, linear, and pure quadratic. We tried all of them and found there is not much difference. Finally, we chose the constant basis function because it is commonly used and takes a little less time. As far as we understand, the coefficients are estimated in the fitting of a GPR model by an optimizer like LBFGS-based quasi-Newton approximation to the Hessian. The algorithm is implemented in MATLAB.

C22

-Pg. 12, In. 286: Please describe in more detail why you are using a kernel function (I assume because then you don't need to actually define the functions "phi(x)").

Yes, exactly. As far as we understand, a kernel function can be used to replace the calculation of inner products or covariance of the outputs of two functions. More details please refer to e.g. Rasmussen and Williams (2006) p.12 and p.14, Duvenaud (2014) Chapter 2, etc.

-Pg. 12, In. 288-290: I don't think these sentences are needed, since in the next paragraph, you thoroughly introduce the ARD-SE kernel.

We introduce the ARD-SE kernel in detail because we have chosen this kernel in our study. But in general, kernel function is one of the hyperparameters which should be chosen according to data. Therefore, we consider it is necessary to give a short overview about kernel functions.

-Pg. 12, In. 291: Why is the ARD-SE kernel chosen? And please provide a reference about this kernel.

We've tried both types of kernels and found out that applying the ARD kernels can obtain much better model performance than applying the kernels with same characteristic length scale, but the results of different ARD kernels, e.g. ARD-SE or ARD exponential, don't show much different. The reference for the ARD-SE kernel is cited in I.298. The same citation has been added to the first mention of the ARD-SE kernel.

-Eq. 21: Please define "D".

Thank you for your suggestion. D is the dimension of predictors and has now been defined in the corresponding text.

-Pg. 12, In. 296: "A relatively large length scale indicates a relatively small

C23

variation along corresponding dimensions in the function" From Eq. 21, it seems more accurate to say that a large length scale relative to the amount of variability in the predictor indicates a smaller variation along the corresponding dimensions. For example, it seems the size of the length scale is only meaningful by itself if all of the predictors have been normalized to the same std. dev. Is this correct?

We think it depends on whether one decides to train the model with standardized data or not. In our study, the training data is standardized using z-scores, i.e. centering and scaling the data by its mean and standard deviation, respectively. This is explained in Sect.5.1 Model Training.

-Eq. 22: Can you state the difference between X and X_* here?

The meaning of $*$ is stated below Eq.22. But if it is not clear, we can modify it.

Also, this equation seems to just be saying that the conditional distribution is normally distributed, so I don't think the right hand side of the equation adds anything. Perhaps it would be less confusing to just explain that the function values are estimated given input parameters X_* by conditioning f on the training parameters and observations, X and y , as well as X_* : $f_* | X, y, X_*$.

We just intended to use Eq.(22) to explain that the predictive equation of GPR is a conditional distribution given the training data and the new input data, and this conditional distribution is normal distributed.

Finally, how are the estimates formed from the resulting distribution? Is the mean value used?

The mean value is the predicted value of the target variable, and the 95% confidence interval is determined by the variance of this distribution.

-Table 2: The lidar weighting function width (e.g., Full Width at Half Maximum) would be a relevant parameter to list in this table.

Thank you for your suggestion. Now, the information about the FWHM of both lidars has been added to the text as well as the table.

C24

-Pg. 15, In. 367: "The threshold for both are 6 m/s and 3 m/s". How are the thresholds used? For example, are these the thresholds in terms of deviation from the mean value of the three-data point window? Also, how is the standard deviation defined here? How many data points are used to calculate the std. dev.?

The range filter works in this way: 1) Calculate the value range within the window (range = max value - min value); 2) If the value range exceeds the preset threshold, this point will be filtered. The standard deviation is the standard deviation of all the values within the window. We noted the explanation about the filter is not well formulated. This will be improved.

-Pg. 18, In. 436: "all the PDFs supported by MATLAB" Is there a particular MATLAB toolbox you are referring to here? Also please provide a reference for MATLAB.

We used a tool called *fitmethis* developed by Francisco de Castro. This tool requires Statistics and Machine Learning Toolbox of MATLAB. We apologize for missing the citation in the text! Citation: Francisco de Castro (2020). *fitmethis* (<https://www.mathworks.com/matlabcentral/fileexchange/40167-fitmethis>), MATLAB Central File Exchange. Retrieved Jan 13, 2020.

-Pg. 20, In. 465: "all the fitted curves of the coherence are grouped together proves it is reasonable to model the wind evolution based on dimensionless frequency". Do you mean that they are grouped together at high frequencies ($f_{dless} > 0.1$)? Additionally, "proves" seems like a strong statement here. Maybe "suggests"?

Yes. And thank you for your suggestion for wording.

-Fig. 9: The caption should refer to the different subplots that are labeled (a-d).

Thank you for pointing out the missing information in the caption. This information has

C25

now been added.

-Pg. 22, In. 486: "all the potential predictors are included...to determine the characteristic length scale". Please describe how the training to find the length scales is performed.

As same as coefficients β , characteristic length scale(s) is estimated in the fitting of a GPR model by an optimizer like LBFGS-based quasi-Newton approximation to the Hessian.

-Pg. 22, In. 489: "Figure 11 illustrates a comparison among the $\log(\sigma_m)$." As mentioned earlier, it doesn't seem fair to compare the σ_m magnitudes unless all of the predictor variables have been normalized to have the same std. dev. (or some other normalization). Is this done?

Yes. The training data is standardized by centering and scaling the data of each predictor by its mean and standard deviation, respectively, which gives the standard scores (also called z-scores) of the predictor data.

-Table 4: Please explain "standard deviation of observed responses" in more detail. It's not clear what the "observed responses" are.

The "observed response" generally means the model response observed from the data. In our case, it refers to the target variables, i.e. the fitted wind evolution model parameters a and b . We will modify the corresponding text to make it clearer.

-Fig. 13: I'm not sure how to interpret this figure. Are there errors in the plots or the legend? The legend lists separate solid lines and dotted lines, but I don't see both in the plots. Do they perfectly overlap?

In this example, yes. We will modify the line colors or the line styles to make the curves distinguishable even though they overlap.

Additionally, the legend says that blue dotted is a fitted case and solid red is the

C26

predicted case. But these two lines are very far apart, which does not support the claim that the fitted and predicted curves are very close.

Different line styles indicate different types of curves: line with dot is fitted curve, normal line is predicted curve, and dashed line is confidence interval. Different colors indicate the results for different separations: blue is for R_1 vs R_2 and red is for R_1 vs R_5 . And there is indeed an error in the color of the predicted curves in the legend. It has now been corrected.

Additionally, what is the significance of the particular period being shown here. Is this one of the periods with the best match between the fitted and predicted coherence? Or is it representative of a typical case?

We did not intend to show an example with the best match. We chose this data block based on two principles: 1) data integrity and 2) representative wind statistics. In this example, the lidar measured mean wind speed is 7.3–7.7 m/s and the lidar measured turbulence intensity is 0.10–0.12, for different range gates. These values appeared frequently in the selected period according to Fig. A1. Hence, we think this example is representative for the data involved in this study. And we decided to choose a data block as an example because we thought neither the coherence itself nor its prediction are intuitive (esp. a and b are predicted respectively but must be combined to give the modeled coherence). We noted a mistake here: the date should be 07 Dec. 2013. Figure 6, 8, and 13 are all plotted with the same data block. The date has now been corrected.

-Pg. 26, In. 556: "R² at least over 0.65." What is the significance of 0.65 as an indication that the prediction accuracy is "satisfactory"?

We have to admit that the wording here is not very appropriate. We will modify the inappropriate expressions in the paper.

-Pg. 26, In. 573: "no obvious relevance between the error and values of any of the predictors is indicated in both figures." I don't quite agree. I think there are some

C27

interesting trends, like in Fig. 14 (b), the RMS prediction error decreases as σ_{u_i} increases. Trends can also be observed in Fig. 15 (b) and (f).

Thank you for your suggestion. That is a very interesting finding. We will think about this part again.

-Pg. 29, In. 612: "capable of achieving a parameterization model with sufficient accuracy for the prediction of wind evolution." This seems like a very strong statement to make. Please provide some more context for this statement. How is "sufficient accuracy" determined?

We noted the wording here is not very appropriate. We will modify this statement to make it more objective.

-Pg. 29, In. 616: "methods to improve the estimation of the coherence and the wind statistics are desired." What are some of the shortcomings of your current approaches that you think could be improved?

For example, if the direction misalignment could be determined in a reliable way by e.g. using a lidar with multiple beams, it might be possible to use a more sophisticated wind evolution model to analytically account for more complicated effects. moreover, methods to improve the accuracy of turbulence intensity or high order wind statistics derived from lidar data would be of great interest (if it is possible).

-Fig. A1-A5: I think if there are appendix figures, they should be in a labeled appendix section.

Thank you for your suggestion. An appendix section has been added.

-Pg. 30, In. 623: Is the current computational time acceptable for real-time applications?

According to our study, it is possible to do real-time prediction but not real-time model training.

C28

-Section 6: If possible, it would be nice to hear your thoughts on whether the chosen coherence formula structure (Eq.5) could be improved. In other words, the paper mostly focuses on how to estimate the a and b parameters, assuming Eq. 5 is the right model. But is this model good enough? For example, Simley and Pao (2015) show that this kind of model did not fit stable atmospheric conditions very well.

We initially also wanted to study the influence of atmospheric stability on wind evolution. However, we found that the stability of the selected period of LidarComplex (where sonic data is available) happened to be mostly neutral. Therefore, unfortunately, we could not get any conclusions related to atmospheric stability.

According to our experience, we think this coherence formula structure is reasonable and can fit the estimated coherence well in most cases as long as noises in the high frequency range (if exist, e.g. the noises caused by motion of the nacelle) can be properly filtered. And we found that in comparison to fitting all coherence measured simultaneously (between different separations) at once (as Simley and Pao (2015) did), fitting the coherence individually could improve the fitting quality because this enable each coherence to find its best-fit parameters.

C29

Response to comments of Mark Kelly

General comments

This work examines evolution of advected turbulence, in terms of spectral coherence, with the motivation of wind turbine control assisted by inflow measured by lidar. The topic is quite relevant to wind energy, and fits well with the journal (WES).

There is some interesting content and potentially useful results, with the use of GPR and Bayesian inference being quite nice.

We would like to thank the referee for the interest in this research and the positive feedback on our methodology.

Unfortunately the paper appears to be somewhat 'unfinished'; perhaps it is also due to the lack of English fluency or preparation time. For example, the abstract has simply copy-pasted a few sentences from the paper, and repeats in a cumbersome way: 'This paper aims to achieve parameterization model for the wind evolution model to predict the wind evolution model parameters'. The paper needs to be proofread by somebody with English fluency, at any rate.

The abstract does not clearly provide an idea of the work done, and while the text has more detail, it is not clear throughout; I am not sure that readers could repeat what has been done.

Thank you for your suggestion. We will improve the presentation of the whole paper, especially the abstract. The revised manuscript will be proofread before submission.

More importantly, there are inconsistencies that have not been considered, and should be addressed/rectified; perhaps most significant are the form itself chosen for coherence (see derivations below in Specific comments), and the use of Taylor's hypothesis for some (but not other) parts of the model/parameterization.

C30

Thank you for your suggestion. Please find below our responses to the related comments.

There are a number of details and also explanations which are missing, but which could hopefully be included, to make the work publishable. The results/performance are a bit overstated (in English), but this is not needed, as the numerical results presented tell the story less subjectively—and are good to share with the wind energy community, provided that they are given with sufficient detail, replicability, and consideration.

Thank you for your suggestion. We will include the missing information and modify the wording of the paper as suggested by the reviewer.

Specific comments

Line 15 and elsewhere later in the paper: while the authors define ‘wind evolution’ as squared coherence, they imprecisely define such (e.g. “Coherence is a dimensionless statistic in the frequency domain”).

When we reviewed the relevant literature (see the following references as examples), we noted that squared coherence is commonly used instead of magnitude coherence (Although it is not clearly stated as "squared coherence" in the text, but "coherence", the formulas show squared coherence.). Hence, we decided to follow their definition to keep our work consistent with the previous studies. But we agree that the expression "Coherence is a dimensionless statistic..." is not precise enough, and thus we’ve now rephrased the corresponding text.

Davenport, A. G. (1961). The spectrum of horizontal gustiness near the ground in high winds. *Quarterly Journal of the Royal Meteorological Society*, 87(372), 194–211. <https://doi.org/10.1002/qj.49708737208>

Panofsky, H. A., & McCormick, R. A. (1954). Properties of spectra of atmospheric turbulence at 100 metres. *Quarterly Journal of the Royal Meteorological Society*, 80(346), 546–564. <https://doi.org/10.1002/qj.49708034604>

C31

Specifically, this should be written as temporal coherence (time shift, frequency domain), in contrast with spatial (wavenumber spectra) coherence. Further, coherence does not describe the “correlation between two signals”, but rather the correlation between spectral components of two signals.

Thank you for your suggestion. This paragraph has now been rephrased accordingly.

Do lines 19–20 not imply that use of Taylor’s hypothesis means ignoring wind evolution? This may be relevant, for consistency later (line 235).

Yes. We think that the degree of wind evolution should depend on the evolution time, so Taylor’s hypothesis should be valid when the evolution time approaches to zero. We noted that the text in l.232-235 is not well formulated, and thus we’ve now rephrased that part.

Line 35: The statement “dependence of coherence on separation and atmospheric stability was not adequately researched” lacks reference and/or explanation. It was not adequate, according to whom, or how?

The content in l.29-35 is a brief introduction of the work of Panofsky and Mizuno (1975). This statement is specifically for that study. To avoid misunderstanding, we’ve added "in that study" at the end of the sentence.

Line 36–38: You write “The longitudinal coherence differs from the lateral and vertical coherence because the former measures the correlation with respect to time lag while the latter with respect to spatial separation.” This is not correct: longitudinal, lateral, and vertical coherences all depend on f , based on integration over time lags; they are depending on spatial separations in the respective directions. For the longitudinal coherence to give spectral correlations ‘with respect to time lag’ Δt , then the longitudinal separation is related to Δt in some way, though you have stated before this point that you are not using Taylor’s hypothesis.

We have to admit that this sentence is not well formulated, and maybe the use of "with

C32

respect to" is not proper here. Of course, coherence is a function of frequency when it is estimated from time series data. What we wanted to express is that the lateral and vertical coherence depends on spatial separation in the respective directions, while the longitudinal coherence is coupled with the time-dependent variation of turbulence because the evolving eddies are moving in the longitudinal direction with the mean flow. So, the longitudinal separation is related to the travel time (corresponding to the evolution time) Δt . And we think this is independent of application of Taylor's hypothesis, but Taylor's hypothesis provides an approximation (x/U) for the travel time.

Line 112/equation 2: how is this a function of frequency (f)? I.e., include the f-dependence on LHS, and also within τ on RHS.

Thank you for your suggestion. We've now included the f-dependence on LHS. But regarding the f-dependence of τ on RHS, because it is derived from Eq.(3) and Eq.(4), which is "unknown" for Eq.(2), we would prefer not to include it in Eq.(2).

There appears to be incompatibility between Eqs. 3–4 and Eq. 5; in particular (5) is missing σ and U .

We have to admit that here we may have written too briefly. The sentence "Combining Eq. (2)–(4) and introducing the second parameter in the model, as inspired by Simley's model (2015a)..." means that combining Eq. (2)–(4) gives the formula like Pielke & Panofsky (1970), and then we imitate the formula of Simley's model (2015a) to introduce a second parameter in the model. We initially wanted to avoid introducing too many formulas for brevity. But if that would cause confusion, we will add related formulas and explanations.

Further, in Section 4.2 and Fig. 10, you analyze the behavior of a with $\Delta t_{\text{maxcorr}}$ (stating $a \propto \Delta t_{\text{maxcorr}}^{-0.49}$), but do not consider that the equations already imply a Δt dependence. That is, eqns. 2 and 5 give

$$C\Delta t/\tau = \sqrt{(af\Delta t)^2 + b^2}; \quad (2)$$

C33

however (3)–(4) with (2) give

$$(CI_u f\Delta t)^2 = (af\Delta t)^2 + b^2 \quad (3)$$

where the turbulence intensity is defined $I_u \equiv \sigma/U$. Thus one sees that

$$a = \sqrt{(CI_u)^2 - (b/f_{\text{dless}})^2} = CI_u \sqrt{1 - \left(\frac{b}{C\Delta t/\tau}\right)^2}. \quad (4)$$

In the limit of high dimensionless frequency or large turbulence intensity, i.e. without the offset b , then $a \rightarrow CI_u$ like Pielke & Panofsky (1970). But in the limit of small turbulence intensity or low dimensionless frequency (small $f\Delta t$), i.e. a large offset, then we see a become imaginary, implying a (nonphysical) coherence oscillating with Δt or Δx .

Thank you for pointing out this interesting question. In fact, we did notice it and have examined if the Δt dependence of a is due to the Δt introduced in f_{dless} . As presented in the paper, $a \propto \Delta t_{\text{maxcorr}}^{-0.49}$. If this dependence originally did not exist but entirely comes from Δt in f_{dless} , the exponent of t_{maxcorr} should be around -1, and it could be canceled out with the Δt in f_{dless} . Moreover, we also tried to fit the estimated coherence to the coherence model dependent of frequency or wavenumber (without introducing Δt in the formula), and we found that $a' \propto \Delta t_{\text{maxcorr}}^{0.51}$ (here use a' to distinguish from the former a). Hence, we think that a (or a') originally has a Δt dependence, and the Δt in f_{dless} just changes the exponent of Δt .

The form (5) is the same as that of (4) in Simley & Pao (2015), with d_l in the latter replaced here by $U\Delta t$, and b here replacing their abd_l ; this should be noted, and the text is not quite clear nor correct following your Eqn. 5. You do note the reason for keeping Δt (instead of using $\Delta t = d_l/U$), but why did you drop the spatial separation dependence (d_l) from the 'b' part of the Simley Pao (2015) expression? From your logic for the 'a' term, then instead of just b you would have bd_l (but not $bU\Delta t$).

The reasons for using b to replace their $ab'd_l$ (use b' to distinguish from our b) are: 1)

C34

In terms of curve fitting, ab' is essentially the fitted term, and thus b' shows a strong dependence on a , which is generally undesirable for machine learning methods. 2) Using $ab'd_l$ (or $b'd_l$) implies that the unit of b' is m^{-1} , while a is dimensionless. We wanted to make both dimensionless to keep consistent. 3) We did try $b''d_l$, but we found that d_l is still an important predictor for b'' , which indicates b'' still depending on d_l . Then, it is not necessary to assume $b''d_l$, but simply use b and take d_l as a predictor.

In lines 145-9: your text is a bit imprecise here – the sonic anemometer has a measuring volume as well (not a point), it is just much smaller than the lidar's. Also, among the reasons why longitudinal coherence from lidar deviates from that calculated via sonic anemometers, one key possibility is missing: the validity of Taylor's hypothesis.

Thank you for your suggestion. We will modify the text accordingly.

Line 153: what do you mean by "complete coherence curve"?

In that context, a "complete coherence curve" means the coherence can more or less cover the range from the highest coherence (e.g. 0.9 – 1.0, sometimes could be lower depending on spatial separation) to lowest coherence (e.g. 0 – 0.1). We've now improved the expression to make it clearer.

Lines 157–164: please include references, as this is not original.

Thank you for your suggestion. The relevant reference has now been cited to the corresponding text.

Line 184: neglect of the spatial averaging effect and $w(x)$ in $\gamma_{i,j}^2$ also demands use of Taylor's hypothesis. This should be noted (along with its potential inconsistency).

Here, Taylor's hypothesis is applied within the lidar probe volume. The lidar probe volume is resulted from the length of laser pulses, with typical length in the order

C35

of magnitude of 10^{-7} s (e.g. 150-400 ns depending of devices). Considering the light speed, this temporal length corresponds to a spatial length in the order of magnitude of $10^1 - 10^2$ m. Within this range, the exact locations from which the signals are backscattered can not be distinguished. Because wind evolution depends on the evolution time of turbulence (see equation (2) in the paper). In this case, the corresponding evolution time is in the order of magnitude of 10^{-7} s. Therefore, wind evolution can be neglected within the probe volume. Referee 2 asked a similar question. Please find in p.11 the discussion if you are interested in more details. And we noted a mistake: in fact, it is not necessary to specifically assume $t = x/U$. There must be a correspondence between t and x as long as wind flows in x direction.

Lines 193–5: the sentence "To retrieve the longitudinal coherence in this case, the above discussed spatial averaging effect must be coupled to a specific turbulence model (Schlipf, 2015; Mann et al., 2009), and thus the wind evolution model is included in the final model implicitly" does not quite make sense. Could you clarify?

Schlipf et al. (2015) suggested an approach to consider different effects of lidars when detecting wind evolution. Here, we briefly mention the explicit expression of the horizontal coherence deduced in that study, based on the assumption of lidar point-measurement for simplification:

$$\gamma_{ij,losP} = \frac{\cos^2(\alpha_H)\gamma_{ij,ux}\gamma_{ij,uy}S_{ii,u}}{\cos^2(\alpha_H)S_{ii,u} + \sin^2(\alpha_H)S_{ii,v}}, \quad (5)$$

where $\gamma_{ij,losP}$ is the horizontal coherence of lidar point-measurements, $\gamma_{ij,ux}$ and $\gamma_{ij,uy}$ are the longitudinal and lateral coherence of the u -component, $S_{ii,u}$ and $S_{ii,v}$ are the auto-spectra of u and v components, α_H is the misalignment angle. From this equation, one can see the determination of the longitudinal coherence $\gamma_{ij,ux}$ is only possible given a specific turbulence model (knowing $S_{ii,u}$, $S_{ii,v}$ and $\gamma_{ij,uy}$) and knowing the misalignment angle α_H .

The volume averaging effect of lidar is then taken into account with a Riemann sum

C36

based on the theoretical consideration for the case of lidar point-measurement, and thus the equation is too complex to be explicitly expressed.

More details please see:

Schlipf, D., Haizmann, F., Cosack, N., Siebers, T., and Cheng, P. W.: Detection of Wind Evolution and Lidar Trajectory Optimization for Lidar-Assisted Wind Turbine Control, *Meteorologische Zeitschrift*, 24, 565–579, <https://doi.org/10.1127/metz/2015/0634>, 2015.

Line 217: by “its definition”, do you mean the definition analogous to (13), where the timelags is replaced by spatial separation r ?

Yes, exactly.

Lines 217–19: If you say $L_{int} = T_{int}U$, then aren't you just using U as a potential predictor somehow?

Yes. U itself is also included as a predictor.

Also, isn't this inconsistent with the previous section, where you state that Taylor's hypothesis is to be avoided? Or, is Taylor's hypothesis avoided only for certain aspects? Please clarify.

Unfortunately, for calculation of L_{int} from measured data, we have no alternatives except this approximation. But in this part, we only discuss in general what could be the possible predictors. A prediction selection is done to select proper predictors (discussion see Sect.5.2).

Line 223: The statement “atmospheric stability represents a global effect of the boundary layer on the wind field” is not quite correct. From what you measure, or via M-O similarity, it is a ‘global’ effect from the surface, and potentially only through part of the ABL (sometimes not even above the surface layer in stable conditions).

Thank you for your suggestion. We will modify the text accordingly.

C37

Line 234: “So as the travel time Δt .” is not a sentence. What are you trying to convey here?

Thank you for pointing out this mistake in writing. We've now modified this paragraph to make it clearer.

Line 235: So you are meaning that distance d is used instead as a predictor.

Yes, exactly.

Line 250: by “performs the best”, perhaps you should use ‘performs well’ unless you explain what ‘best’ means (i.e. what other models).

Thank you for your suggestion. In a preliminary study, we explored different machine learning algorithms on a simple level, including stepwise linear regression, regression tree, support vector machine regression, and Gaussian process regression, and we found GPR performs the best among these methods. Now, we've added these details and moved this part to the introduction.

Line 257: The phrase “underlying functions of the data” is not clear. Do you mean behavior conditioned on other variables, or relation to other variables?

Here, we mean “one needs to initially guess what type of function(s) could exist among the relevant variables before choosing a specific regression model.” We will rephrase this part to make it clearer.

Table 1/ line 255+ : Is it even possible to use the fourth or even third moment, given the large sampling uncertainty involved for higher-order moments? Please see and reference e.g. Lenschow, Mann Kristensen (1994) and Ch.2 of Wyngaard's text-book (2010), to understand and defend use of μ_4 let alone μ_3 .

Thank you for the recommended references. We agree that the third and fourth moment determined from measured data would contain a large uncertainty. But our approach is first to find all variables which can be obtained from measured data and

C38

then to use feature selection to select suitable variables as predictors for the GPR models. Feature selection can detect the statistical correlation between predictors and the target variable. Although some variables could contain uncertainties, these variables could still be useful for the prediction of GPR models as long as they have a strong statistical correlation with the target variable. From this perspective, machine learning algorithms generally have an error tolerance for data (of course, the more accurate the data is, the more accurate the prediction could be), which is also one of the advantages of machine learning.

Line 284: To be clear and consistent, can you not specify that β is a weight, and the 'basis function' $h(x)$ maps the means into the new space?

β can also be understood as the weight vector of $h(x)$. But we've defined w as a weight vector before, we wanted to avoid using the same word in case reader might confuse these two different processes: $\mathbf{h}(x)^\top \beta$ is used to model the mean function $m(x)$ of $g(x)$, and $g(x) \sim \mathcal{GP}(m(x), k(x, x'))$ ($k(x, x')$ is a kernel); $\phi(x)^\top w$ is used to find the linear model of $f(x)$ in a higher dimensional space. Theoretically, for any function $f(x)$, it is always possible to find a linear model equivalent to $f(x)$ in a higher dimensional space. For example, for a quadratic function

$$f(x) = ax^2 + bx + c, \quad (6)$$

if define $p = x^2$, $q = x$ and $r = 1$, then

$$f(x) = g(p, q, r) = ap + bq + cr, \quad (7)$$

which is a linear model of $f(x)$ in the three-dimensional space of (p, q, r) . In the algorithm of GPR, $\phi(x)$ is not explicitly defined. The mapping is done through a kernel, which is the so-called kernel trick.

What is meant by 'Basis function is one of the hyperparameters'? I.e., how is a function a parameter, or is $h(x)$ already assumed to have some form, possibly related

C39

to the $\varphi(x)$ forms?

A hyperparameter, different from a parameter, is not necessarily a value (or values), but more like a setting adjusting model behavior (specifically related to machine learning). There are four types of basis functions provided in MATLAB: empty (assuming no basis function), constant, linear, and pure quadratic.

lines 289–296: σ_m is not a 'length' in the physical sense; it has units of whatever x_m has. Thus it is a characteristic magnitude for the predictor having index m .

In the context of machine learning, this term is defined as "characteristic length scale". Please see references like:

Rasmussen, C. E. and Williams, C. K. I.: Gaussian processes for machine learning, Adaptive computation and machine learning, MIT, Cambridge, Mass. and London, 2006.

Duvenaud, D.: Automatic model construction with Gaussian processes, Apollo - university of cambridge repository, <https://doi.org/10.17863/CAM.14087>, 2014.

lines 299–302: To be explicit, the RHS of (22) does not contain a 'conditioning bar'. I.e. to help the reader and match the text, show in the math that the joint Gaussian prior is conditioned on X_* (is the eqn. correct?). Most readers will not have read Duvenaud's PhD thesis, so it is useful to help them understand.

We just intended to use Eq.(22) to explain that the predictive equation of GPR is a conditional distribution given the training data and the new input data, and this conditional distribution is normal distributed. We did not include the equations of \bar{f}_* and $cov(f_*)$ because both are very complicated. We don't think it is necessary for this paper to go so deep into mathematics.

Lines 307–311: why $k=5$? If it is due to needing a large enough sample for verification, then this should be stated.

Theoretically, k can be any integer between two and the number of observations

C40

(This is a special case which is called "Leave-one-out"). When k is very small, the sample size of training data ($\frac{k-1}{k}$ of the total observations) could be not large enough. However, the training process must be repeated k times. So, when k is too large, the training could take very long time. Therefore, $k = 5-10$ is common used in machine learning. We will add this explanation in the corresponding text.

Line 335: was the lidar on the nacelle, at what height?

At 95 m. We have also noted some information about the measurements is only listed in Table 2 but missing in the text. We will improve this part.

Line 350: please be more clear and specific, and also include references.

The research project ParkCast is an ongoing project led by our institute (Stuttgart Wind Energy, University of Stuttgart). Because no publications related to this project have been published so far, we cannot cite any references. However, we have communicated the project-related information with our colleague in charge of this project to ensure its correctness.

Line 362–366: Why are two different filtering types used?

Because the two lidars are different, and one of them is a long range lidar (the max range was set as 990 m for the data used in this paper). For a long range lidar, the backscattered signals from distant range gates could be very weak, and thus the CNR values could be low although the measured wind speed is plausible. In this case, filtering the data based on CNR values is not a good idea. Würth et al. (2018) suggested an approach to filter the data with a range filter, which can keep more valid data compared to the CNR filter.

Würth, I., Ellinghaus, S., Wigger, M., Niemeier, M. J., Clifton, A., and Cheng, P. W.: Forecasting wind ramps: Can long-range lidar increase accuracy?, Journal of Physics: Conference Series, 1102, 012 013, <https://doi.org/10.1088/1742-6596/1102/1/012013>, 2018

C41

State why -24dB for CNR; include reference.

We determined the threshold for CNR by checking the plot of CNR values and wind speed. The CNR threshold could be various under different measurement conditions.

Line 375: Using "C2N" to symbolically write 'N choose 2' unique pairs, is not standard practice. You can write $\binom{N}{2}$ or equivalently $N(N-1)/2$.

Thank you for your suggestion. This has now been corrected.

Line 479–480: need citation for Levenberg-Marquardt algorithm.

Thank you for your suggestion. The corresponding citations have now been added to the text here.

Fig.9 caption: mention which plots belong to which campaign.

Thank you for pointing out the missing information. The campaign names have been added to the caption.

Fig.11 / §5.1 : why not plot σ_m^{-2} ? This is what is actually used in the ARD-SE kernel shown in eqn.21, and its behavior more clearly demonstrates relevance.

We can also plot σ_m^{-2} , but in principle, it will not be different because $\log(\sigma_m^{-2}) = -2\log(\sigma_m)$. The plots will just be flipped up side down, with the y-axis (in log) scaled by two. The benefit of using $\log(\sigma_m)$ is that it shows the order of magnitude of σ_m directly.

Line 489–90: you state "predictors are selected according to different preset limits of the $\log(\sigma_m)$ considering different cases of application or data availability", but what are these preset limits?

The limits are listed in Table 5 and discussed in Sect.5.2. We will modify the text to make it clearer.

C42

Line 494 / Table 4: how/why did you choose the initial $\sigma_m = 10$?

The initial values of σ_m are randomly set. They will be estimated from training data and the GPR algorithm just need some initial values to start the training process.

Line 509: I am not sure that R2 of 0.65 is "satisfactory"; perhaps this could be just written in terms of R2 and RMSE without the subjective claim.

Thank you for your suggestion for wording. We will modify it (and other similar expressions).

Also, "all situations" is not quite consistent with just the recommended cases (i.e. it implies all cases).

"All situations" here refer to I.501: "two different situations of data availability are considered: only using variables calculated with lidar data as predictors (in both of LidarComplex and ParkCast available) and only using variables calculated with sonic data (only in LidarComplex available)." We intended to distinguish "situation" and "case", and to use "situation" to indicate different data availability. But maybe "situation" is not a suitable word. We will consider the wording here again.

Lines 519–520: This is a good point, and it would be useful to repeat this earlier, when introducing the potential predictor variables because some of them appear redundant.

It is mentioned in the penultimate paragraph in Sect.2.5, but maybe it is not clear enough. We will indicate this point more clearly.

Lines 524–527: It appears that you are conflating two things here, one of which you are missing—the applicability of Taylor's hypothesis will also affect L compared to T via U , whereas this is not the case for the usage of U to 'convert' σ_u to I_T .

What we wanted to discuss here is the possible difference for the model between using the variables directly acquired from measured data like U and σ and the variables derived from the other variables like $I_T = \sigma/U$. In principle, I_T can be regarded as a

C43

function of σ and U , and thus it is probably "useless" for the model. But we agree that we missed the point — the approximation of L by TU assuming Taylor's hypothesis is less accurate and thus probably less preferred by the model. Thank you for your suggestion.

Lines 532–534: perhaps μ_3 or μ_4 could help prediction; but to be responsible, one needs to mention that [1] the uncertainty in these quantities are very large (Lenschow et al. 1994 reference), and [2] lidar may not be able to consistently measure these. Further, these higher-order moments are likely more affected by your filtering.

Thank you for your suggestion. That is a very good point.

Lines 537–539: this is likely due to the implicit co-dependence I derived above, i.e. a is a function of CIT and b/fd less. Your finding confirms also the need/utility to consider the behavior of the parameters involved.

As mentioned above in the response to the comment on Eq.(5), we did consider this issue. The ideal case is a and b are completely uncorrelated, but the model form determines that the correlation between a and b cannot be completely eliminated. Indeed, we have reduced their dependence by adapting the form of the offset parameter b .

Lines 542–554: what about cross-comparison using the sonic? Were the wind directions such that the sonic (at 270m upstream) could be compared to the lidar (e.g. at 163.5m upstream)?

We are not quite sure what kind of comparison this "cross-comparison" refers to. We assume that it means the coherence estimated from the sonic data and the data measured at the farthest range gate of the lidar when the wind direction is aligned with the line between the met mast and the wind turbine. Firstly, to estimate coherence, this wind direction must exist long enough, which is less likely to happen in practice. Secondly, because the sampling rate of the ultrasonic anemometer is much higher than the lidar, the sampling rate of the sonic data must be artificially reduced, by e.g.

C44

averaging, to match the sampling rate of the lidar data. But this leads to the fact that estimating coherence between the sonic data and the lidar data does not bring more benefits than estimating coherence between lidar data measured at different range gates.

Line 555-6: isn't this 'satisfactory' $R^2 \geq 0.65$ only true for certain cases and variables?

This statement is specific for the cases using lidar data. Although we discussed different variable combinations and compared the performance of the corresponding models, all the variables are essentially derived from line-of-sight wind speed. In other words, as long as there is a lidar measuring wind speed properly, one would be able to derive all the discussed variables. The question is which of them are necessary. In general, we want as many as necessary and as few as possible.

Figure 14 / Line 560 and afterward: these plots do not responsibly/transparantly show prediction error, as they don't give an idea of the magnitude of a . You should plot percentage error or similar; given that a can be small depending on band IT (as derived above), the plotted differences in a might be relatively significant.

We have been thinking for a long time if it would be better to show relative error or absolute error. Our concern is that values of a and b are very abstract and completely not intuitive. In fact, the shape and position of the predicted coherence determined by both parameters together is the final prediction goal. And the prediction error is the shift of the predicted curve from its estimated curve due to the error of a and b .

Assume a prediction error for a is 0.5. Its relative error will be 50% given $a = 1$ and 25% given $a = 2$. But is the difference between the curve of $a = 1$ and $a = 1.5$ somehow related to 50% of something? Or is the difference between the curve of $a = 1$ and $a = 1.5$ somehow as twice large as the difference between the curve of $a = 2$ and $a = 2.5$? It is not the case. In the end, we chose to show the absolute error because knowing the absolute error, one could more or less imagine how much

C45

the predicted curve would be shifted due to the error by observing Fig.3, which is not possible for showing the relative error.

Moreover, we are considering to plot two additional curves on the plot of example predicted curves (Fig.13) to indicate the range of predicted curves due to the RMSE of both parameters. That could hopefully give an intuitive feeling for the prediction error of the coherence curve.

Lines 574–5: if the whiskers are large because of sample size, then why not (also) account for this via \sqrt{n} ?

Here, what we wanted to express is that large whiskers indicate large variances of predicted errors. The reason for that could be insufficient training data in the corresponding value range resulting in less accurate predictions for that value range by the model. We noted the expression here is not clear enough and will improve it.

Line 576: The claim "it is proven that the Gaussian process regression is capable of achieving an accurate parameterization model" is an overstatement. It is DEMONSTRATED(not proven) that the GPR was able to predict two coherence model parameters with an $R^2 \geq$ in chosen cases (not simply 'accurate').

Thank you for your suggestion for wording. We will rephrase this part.

Technical corrections

There are many English usage/grammatical corrections and suggestions, which are included in the attached annotated PDF-file. I thus only include a sample of them here in this list. The sentence structure and writing is unclear or ambiguous in numerous places; the paper really should be reviewed and edited by somebody with adequate fluency.

Thank you for your suggestion. The revised manuscript will be proofread before sub-

C46

mission.

I list some of the specific corrections below, but since there are >300, after the first few I include the line numbers, which refer to the annotated attached PDF. After page 16 I did not correct much English; this is left to the authors for the next draft.

• Abstract/line 1: One (generally) shouldn't copy sentences from the introduction into the abstract (further, this first sentence is a definition); 'turbulence' should be 'turbulent'; pluralize 'structure'; delete 'of the eddies'; replace 'while the eddies' with e.g. 'as they'; replace 'by the main flow over' with 'through'.

• L.2–4: Remove 'the'; change 'because' to ':'; remove 'only'...see annotated PDF for more details.

• L.5–7: These 2 sentences are quite unwieldy (cumbersome) and also somewhat tautological—particularly for an abstract, also with repeated phrases that need to be reduced/condensed. Please correct the English usage here.

• L.12–13: First sentence can be corrected from "Wind evolution refers to the physical phenomenon that the turbulence structure of the eddies changes over time while the eddies are advected by the main flow over space." to something like 'WIND EVOLUTION' REFERS TO THE PHYSICAL PHENOMENON OF TURBULENCE STRUCTURES (EDDIES) CHANGING OVER TIME, WHILE THE EDDIES ARE ADVECTED THROUGH SPACE BY THE MEAN FLOW.

• L.13-15: change "The mathematical" to 'A common statistical'; delete 'usually'; 'hereinafter for brevity, also' should be 'hereafter'; delete 'two time series data sets of the' and instead add 'measured at two different locations,' after 'velocity'; change 'with certain time shift' to 'calculated over varying time shifts'.

• L.17–19

• L.22

• L.25–27

• L.29

• L.32: run-on sentence; use parentheses as noted

C47

• L.33–34

• L.36–38

• L.40, L.42–44

• L.47–48

• L.52 use of definite articles and parentheses

• L.55–59, L.61

• L.63 delete 'model'

• L.65, 67

• L.73–74

• L.79–81: delete a number of redundant words, add punctuation as noted

• L.122–123

• L.136

• L.151 Capitalize 'Oppenheim', here and elsewhere.

• L.156

• L.169

• L.177: have already introduced U as mean wind speed (delete here).

• Page 8: L.186–7; 189; 196–9; 201–2

• Page 9: L.204; 210; 212; 220–2;

• L.224 and elsewhere: not 'Monin-Obukhov length', just use 'Obukhov length'

• Page 10: L.225–8; 233; 250–7

• Page 11: L.258–9; 261; 264; 269; 272

• Page 12: L.275–8; 288–9; 292; 295–6; 302; 304–5

• Page 13: L.308

• Page 14: L.335; 339; 346–7; 349–353; 363–5

• Page 15: L.367

• Page 20: L.470; 479

• Page 24, Table 5: Taylor is italicized under case 6, but should be Roman font.

• Page 25: L.537; 550

C48

Please also note the supplement to this comment: <https://www.wind-energ-sci-discuss.net/wes-2020-50/wes-2020-50-RC3-supplement.pdf>

We highly appreciate these suggestions for technical corrections. All of the comments have been considered, and corresponding corrections have been made. We'd like to thank the reviewer again for taking valuable time to help us to improve this manuscript. The revised manuscript will be proofread before submission.

Please also note the supplement to this comment:
<https://wes.copernicus.org/preprints/wes-2020-50/wes-2020-50-AC1-supplement.pdf>

Interactive comment on Wind Energ. Sci. Discuss., <https://doi.org/10.5194/wes-2020-50>, 2020.

Parameterization of Wind Evolution using Lidar

Yiyin Chen¹, David Schlipf², and Po Wen Cheng¹

¹Stuttgart Wind Energy (SWE) at Institute of Aircraft Design, University of Stuttgart, Allmandring 5b, 70569 Stuttgart, Germany

²Wind Energy Technology Institute, Flensburg University of Applied Sciences, Kanzleistraße 91-93, 24943 Flensburg, Germany

Correspondence: Yiyin Chen (chen@ifb.uni-stuttgart.de)

Abstract. Wind evolution ~~refers to the change of the turbulence structure of the eddies over time while the eddies are advected by the main flow over space. With the development of the~~, i.e. the evolution of turbulence structures over time, has become an increasingly interesting topic in recent years, mainly due to the development of lidar-assisted wind turbine control, modelling of the wind evolution becomes an interesting topic, because the control system should only react to the changes in the wind field which can be predicted accurately over the distance to avoid harmful and unnecessary control action.

This paper aims to achieve a parameterization model for the wind evolution model to predict the which requires accurate prediction of wind evolution model parameters according to the wind field conditions. For this purpose, a to avoid unnecessary or even harmful control actions. Moreover, 4D stochastic wind field simulations can be made possible by integrating wind evolution into standard 3D simulations, to provide a more realistic simulation environment for this control concept. Motivated by these factors, this research aims to investigate the potential of Gaussian process regression in the parameterization of wind evolution. Wind evolution is commonly quantified using magnitude-squared coherence of wind speed and is estimated with lidar data measured by two nacelle-mounted lidars in this research. A two-parameter wind evolution model suggested in literature was applied modified from a previous study is used to model the wind evolution and the wind evolution was estimated using lidar data estimated coherence. A statistical analysis was done to reveal is done for the wind evolution model parameters determined from the estimated coherence, to provide some insights into the characteristics of wind evolution model parameters. Gaussian process regression was applied to achieve the parameterization model. The results have proven the applicability of models are trained with the wind evolution model parameters and different combinations of wind-field-related variables acquired from the lidars and a meteorological mast. The automatic relevance determination squared exponential kernel function is applied to select suitable variables for the models. The performance of the Gaussian process regression model to predict the wind evolution model parameters with sufficient accuracy models is analyzed with respect to different variable combinations, and the selected variables are discussed to shed light on the correlation between wind evolution and these variables.

1 Introduction

~~Wind evolution~~ Wind evolution refers to the physical phenomenon ~~that the turbulence structure of the eddies changes over time while the eddies are advected by the main flow over space. The mathematical measure of the "change" of turbulence~~

25 ~~structure is usually defined~~of turbulence structures (eddies) changing over time, and is defined, in this study, as magnitude-
squared coherence (~~hereinafter for brevity, also dependent on evolution time. Magnitude-squared coherence (hereafter referred~~
~~to as coherence) between two time series data sets of the turbulent velocity with certain time shift. Coherence is a dimensionless~~
~~statistic in the frequency domain that can describe the correlation between~~is a common statistical measure of turbulence
structure properties (see e.g. Panofsky and McCormick, 1954; Davenport, 1961; Panofsky et al., 1974). In general, coherence
30 ~~describes the correlation between spectral components of~~ two signals or data sets, taking ~~value between 0~~values between zero,
for no correlation, to ~~1.0~~unity, for perfect correlation. ~~Whereas,~~ Because turbulent eddies are advected by the mean flow while
evolving, the longitudinal coherence, i.e. coherence of turbulent velocity at locations separated in the mean direction of the
flow, is used to measure wind evolution in practice (see e.g. Schlipf et al., 2015; Simley and Pao, 2015a). And when estimating
the coherence, the data measured at the downstream location should be shifted by the travel time, corresponding to the evolution
35 ~~time, to match the data measured at the upstream location.~~ Taylor's (1938) hypothesis is a special case that assumes ~~the pattern~~
~~of turbulent motion remains unchanged while the eddies are carried by the main flow(Taylor, 1938)~~all turbulent motions remain
unchanged, while eddies move with the mean flow. In other words, Taylor's hypothesis assumes ~~a perfect correlation~~no wind
evolution, which means the coherence is ~~1.0 over the whole frequency range~~unity for all frequencies. The validity of Taylor's
hypothesis was researched in some studies (see e.g. Willis and Deardorff, 1976; Schlipf et al.). Taylor's hypothesis is widely
40 ~~used in data analysis and wind field modeling for the sake of simplification (see e.g. Kelberlau and Mann, 2019; Veers, 1988).~~

The research on wind evolution dates back to ~~the~~ 1970s. Pielke and Panofsky (1970) attempted to generalize some of the
mathematical ~~description of the~~~~descriptions for~~ horizontal variation of turbulence characteristics. The final goal at that time
was to figure out an empirical model of the ~~four dimensional space-time~~ ~~four-dimensional (space-time)~~ structure of turbu-
lence. In Pielke and Panofsky's ~~work (1970)~~(1970) work, the coherence model suggested by Davenport (1961) to describe
45 the correlation between horizontal wind components at different heights, also known as Davenport Geometric Similarity, was
extended into other wind components and separation directions. ~~They~~ Pielke and Panofsky's (1970) model also followed Dav-
enport's idea to approximate the coherence with a simple exponential function ~~with using~~ a single decay parameter, ~~which is~~
~~known as Pielke and Panofsky's model.~~ The decay parameters were assumed to be ~~some constants~~in that research~~constants~~.
After that, Ropelewski et al. (1973) ~~did a systematic study of~~~~systematically studied~~ the coherence for streamwise and cross-
50 stream wind components with horizontal separations. Based on their theoretical discussion, the decay ~~parameters might be~~
~~functions of roughness and stability instead of constants.~~ Panofsky and Mizuno (1975) ~~continued the study and~~ parameter for
longitudinal separation is supposed to be a function of turbulence intensity, which is a function of roughness length and
Richardson number (J. L. Lumley and H. A. Panofsky, 1964). Extending the study, Panofsky and Mizuno (1975) found that
the ~~relationship~~~~relationships~~ between coherence and other parameters were rather complicated. A model for the decay param-
55 eter was proposed based on ~~the discussion of~~its empirical properties. This ~~decay parameter~~ model involves turbulence intensity
accounting for the influence of terrain roughness, ~~standard deviation of the lateral wind component, lateral~~ integral length scale
~~which shows of the longitudinal wind component (which shows a~~ relationship with Richardson number), separation of ~~the~~
~~two~~ observations, and the angle between the wind direction and the measurement line. This model can be regarded as the first
parameterization ~~model for the of~~ Pielke and Panofsky's ~~model, but it was just developed using (1970) model. However, the~~

60 ~~model was developed using only~~ very few observations ~~measured with meteorological tower~~ taken on meteorological towers, and the dependence of coherence on separation and atmospheric stability was not adequately researched in that study.

It is worth mentioning that ~~wind evolution specifically refers to~~ the longitudinal coherence (~~i.e. coherence in the direction of the flow~~) ~~by definition~~. ~~The longitudinal coherence~~ differs from the lateral and vertical coherence because the former ~~measures the correlation with respect to time lag~~ is coupled with time-dependent variations in turbulence, while the latter ~~with respect to spatial separation~~ measures the decay of correlation due to spatial separations in their respective directions. However, in the above-mentioned studies the longitudinal coherence was not clearly distinguished. Kristensen (1979) proposed that the longitudinal coherence should behave differently and deduced an alternative expression for ~~wind evolution, which is known as it, which we refer to~~ Kristensen's (1979) model. This model ~~is based on the assumption~~ assumes that the coherence can be ~~modelled modeled~~ with the probability that an eddy ~~measured observed~~ at the first ~~observation point will pass the second~~ one point can also be observed at the second point, given that: the eddy has not completely faded out during the travel time; and the eddy has been taken towards the second point.

~~In the recent years, wind~~ Wind evolution has become interesting again because of the new concept of lidar-assisted wind turbine control (see e.g. Schlipf, 2015; Simley and Pao, 2015a; Simley et al., 2018). Lidar ~~is~~ more specifically, Doppler wind lidar ~~is~~ a remote sensing device technology which can be used to measure wind speed in a certain spatial range (Weitkamp, 2005). The main idea of lidar-assisted wind turbine control is to enable a feedforward control of ~~the wind turbine wind turbines~~ by using a nacelle-mounted lidar to measure the approaching wind field at some distance upwind. The control system should ~~only react~~ react only to the changes in the wind field which can be predicted accurately ~~over the distance~~, to avoid harmful and unnecessary control ~~action actions~~. This is made possible by applying an adaptive filter to ~~filter remove~~ the uncorrelated part of the lidar signal. An accurate prediction of the wind evolution will thus benefit the filter design. Moreover, the application of Taylor's hypothesis in the wind field simulation is no longer appropriate for ~~modelling modeling~~ the lidar-assisted control system. To solve this problem, ~~Bossanyi (2013) proposed an approach~~ different approaches (see e.g. Bossanyi, 2013; Laks et al., 2013) have been proposed to integrate the wind evolution model ~~in the Veer's~~ within the wind field simulation method ~~(?) of Veers (1988)~~, to make it possible to simulate a four-dimensional wind field.

Some attempts were made to ~~improve the modelling~~ further promote the modeling of wind evolution. Schlipf et al. (2015) suggested an approach to determine the decay parameter in ~~the~~ Pielke and Panofsky's ~~model with measuring data of a nacelle-based lidar~~. ~~Simley and Pao (2015a) (1970) model with data measured by a nacelle-mounted lidar, taking into account the influence of lidar measurement on coherence~~. However, the limitation of this study is that only four one-hour data blocks were examined. Simley and Pao (2015b) attempted to validate the models of Pielke and Panofsky ~~'s model and the Kristensen 's model (1970) and Kristensen (1979)~~ with data from LES wind fields ~~and found out that the both models cannot~~, but found that neither model can always correctly model the coherence as frequency approaches zero. To improve this issue, Simley and Pao (2015a) ~~adapted the~~ tried to apply the coherence model for transverse and vertical separations suggested by Thresher et al. (1981) to the longitudinal coherence. This model has a form similar to Pielke and Panofsky's ~~model by introducing a new parameter to adjust the vertical intercept of the coherence model so that it can capture the decrease of the coherence at very low frequency~~ (1970) model but includes an additional parameter to allow coherence less than unity at

95 very low frequency. Davoust and von Terzi (2016) examined Simley and Pao's (2015a) model with data from nacelle-mounted lidars on three sites. To enable a direct comparison with Simley and Pao's (2015a) work, a correction method was applied to compensate the influence of lidar measurement on coherence. However, the linear dependence of the decay parameter on turbulence intensity suggested by Simley and Pao (2015a) was not clearly observed. The relationship between the offset parameter and integral length scale shows a good match with that suggested in Simley and Pao's (2015a) work, but the
100 agreement decreases after the correction of coherence. At the same time, de Maré and Mann (2016) developed a four-dimensional model to describe the space-time structure of turbulence by combining the Mann (1994) spectral velocity tensor and the Kristensen's (1979) longitudinal coherence model.

Motivated by the above-mentioned research, this study aims to ~~further analyze how wind evolution is influenced by the wind field conditions and to achieve a parameterization model for an empirical~~ achieve parameterization models for a wind evolution
105 model to predict the modified from Simley and Pao's (2015a) model. In addition, it is desired to gain some insights into the complex relationships between wind evolution and wind-field-related variables such as wind ~~evolution according to relevant conditions, such as the wind~~ statistics, atmospheric stability, and relative ~~position of the~~ positions of measurement points. ~~This~~ For these purposes, a previous study (Chen, 2019) was done to explore different supervised machine learning algorithms on a simple level, including stepwise linear regression (see e.g. Hocking, 1976), regression tree (see e.g. Breiman et al., 1984),
110 support vector regression (see e.g. Vapnik, 1995), and Gaussian process regression (see e.g. Rasmussen and Williams, 2006). It was found that Gaussian process regression, overall, performs the best for prediction of wind evolution model parameters, and thus its potential is further analyzed in this study with more extensive data.

This research is mainly done using lidar measurement because lidar can provide large ~~amount~~ amounts of spatially separated measuring points simultaneously, which is ~~a~~ of great advantage for studying the ~~dependeney~~ dependence of wind evolution on
115 separation ~~compared to in comparison to data from a~~ meteorological tower. ~~Some lidar~~ Lidar data from two measurement campaigns undertaken in different terrain types are available. ~~If any data of the meteorological tower is also available in the~~ corresponding ~~In one of the~~ measurement campaigns, ~~it is~~ data taken on a meteorological tower is also involved in the analysis to provide a comparison.

~~This study is carried out as follows: After the preliminary data processing, wind evolution is estimated with lidar data and fitted to a wind evolution model to obtain the model parameters. The distributions of the obtained wind evolution model parameters are analyzed to figure out some common characteristics. Then, the parameterization model is achieved by training a Gaussian Process Regression (GPR) model. The Automatic Relevance Determination Squared Exponential kernel (ARD-SE kernel) is applied to estimate the relative importance of the different wind field condition variables and to select the suitable input variables for the GPR model. Finally, the performance of the parameterization model is evaluated with a 5-fold cross-validation.~~
125

The present paper is organized as follows: Section 2 briefly explains the theoretical basis of wind evolution and ~~the concept of the parameterization model~~ its prediction concept as well as the principles of the ~~applied methods~~ methods applied in this work; Section 3 introduces the ~~involved measurements and the procedure of the~~ measurement campaigns and the data processing; Section 4 presents the results of the statistical analysis of the wind evolution model parameters; Section 5 illustrates the process

130 of ~~the~~ model training and the evaluation of the ~~obtained~~ parameterization models; Section 6 summarizes the ~~presented results~~
~~again~~ results and gives the conclusions and ~~outlook of this study~~ an outlook.

2 Methodology

This section first explains the mathematical ~~definition-expression~~ of wind evolution in Sect. 2.1. Then, ~~the model concept for~~
~~the prediction of the wind evolution and the corresponding workflow of this study~~ our concept of wind evolution prediction and
135 a corresponding workflow are presented in Sect. 2.2. After that, the wind evolution model applied in this work is introduced in
Sect. 2.3. Finally, the details of the workflow are introduced and discussed in Sect. ~~2.3~~ 2.4–2.7.

2.1 Wind Evolution

As mentioned in the introduction, wind evolution is mathematically defined as the magnitude-squared coherence between two
wind speed signals i and j measured at two points separated in the longitudinal direction, with i for the signal measured at the
140 upstream point and j ~~for~~ at the downstream point:

$$\gamma_{ij}^2(f) = \frac{|S_{ij}(f)|^2}{S_{ii}(f)S_{jj}(f)}, \quad (1)$$

where $S_{ii}(f)$ and $S_{jj}(f)$ represent the power-spectral densities (PSDs) of signals i and j , respectively, and $S_{ij}(f)$ represents
the cross-spectral density between i and j . It must be emphasized that the coherence corresponds to a lagged correlation, which
means the signal j should be shifted by the travel time Δt ~~that~~ after which the signal i is expected to arrive at the downstream
145 point for calculation of the coherence.

2.2 Concept and Workflow

~~It is aimed to predict the wind evolution~~ A supervised learning algorithm aims to find the mapping function from predictors
(i.e. input variables) to a target (i.e output variable) through known data about the predictors and the target without relying on
a predefined equation as a model. The key to using supervised learning is to identify suitable predictors and targets, which is
150 in fact a process of abstracting and condensing information.

In this study, we aim to develop a predictive model for wind evolution of the longitudinal wind component. It is worth
noting the different meanings of wind evolution and wind evolution model. Wind evolution, i.e. the coherence, ~~but it is not~~
~~possible to predict every point of the coherence curve~~ estimated from measured data in practice, is not predictable because the
estimated coherence consists of approximately infinite data points. Therefore, a model with a limited number of parameters
155 is needed to ~~describe the coherence with limited parameters, which then can be predicted by a parameterization model, i.e.~~
~~a surrogate model. That is the~~ approximate the estimated coherence; this is a wind evolution model. ~~Figure ?? shows the~~
~~prediction concept of this work. The idea is first to predict the model parameters of the wind evolution model according to the~~
~~wind field conditions by a parameterization model. Then, the coherence can be reconstructed with the wind evolution model~~
~~using the predicted model parameters.~~ From the perspective of machine learning, using a wind evolution model is essentially

160 condensing the information in the estimated coherence into several model parameters which are predictable. These model parameters are targets of predictive models, and thus the predictive model is deemed a *parameterization model* in this study.

~~Concept of predicting the wind evolution by a parameterization model:~~ wind-field-related variables such as wind statistics, atmospheric stability, and relative positions of measurement points are considered as *potential* predictors, based on the theoretical and experimental studies mentioned in the introduction. A discussion about the potential predictors is provided in Sect. 2.5.
165 Further analysis needs to be done to determine which of the potential predictors should be selected for model training, i.e. *feature selection*. The principle of feature selection is to figure out which variables provide the best predictive power (accounting for most of the variation in the target values) and, ideally, these variables should be independent from each other to prevent over-fitting in model training. To investigate the necessary predictors under different data availability, different combinations of predictors are discussed in Sect. 5.

170 ~~Based on this concept, the study is carried out as follows (see Fig. ??)~~ Figure 1 illustrates our concept and workflow of wind evolution prediction. For model training, the essential steps are determination of observed values of predictors and targets from measured data and training parameterization models using a machine learning algorithm, more specifically: 1) ~~estimation of to estimate~~ the coherence using lidar data; 2) ~~determination of the to determine~~ the observed target values, i.e. the wind evolution model parameters, by fitting the estimated coherence to ~~the a~~ wind evolution model; 3) ~~calculation of the potential predictors from the measured data~~; to calculate observed predictor values from measured data (mainly lidar data; sonic data could be used if available); 4) ~~training the Gaussian process regression model. More details are introduced in Sect. 2.3–2.7.~~ to train parameterization models using a machine learning algorithm. The prediction process goes in the opposite direction: Firstly, the wind evolution model parameters are predicted by the trained parameterization models using new predictor values calculated from new measured data; then, the predicted coherence is reconstructed by the wind evolution model using the predicted model parameters.
175
180

~~Workflow:~~

~~2.3~~ **Wind Evolution Model**

~~As explained above, it is very important to choose a~~ To demonstrate our concept and workflow: Sect.2.3 explains the wind evolution model that can match the trend of the coherence with only several simple parameters. A two-parameter wind evolution model, similar to the Simley's model (2015a), is used
185 used in this study; Sect. 2.4 discusses special issues about coherence estimation using lidar data; Sect. 2.5 discusses the potential predictors of the parameterization models; Sect. 2.6 and Sect. 2.7 briefly introduce the principle of Gaussian process regression (the machine learning algorithm applied in this study) and the method of model validation, respectively; Section 3.2 shows the fitting process of the estimated coherence in detail; Section 5 demonstrates the training of parameterization models, predictor selection, and model validation in the respective subsections.

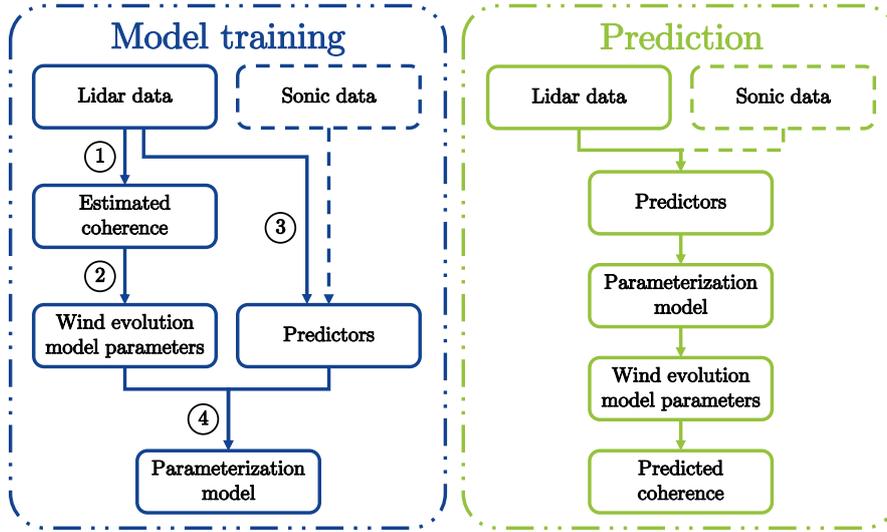


Figure 1. Concept and workflow of wind evolution prediction. The workflow of model training is: 1 – Estimation of coherence using lidar data; 2 – Determination of wind evolution model parameters by fitting the estimated coherence to a wind evolution model; 3 – Calculation of potential predictors from measured data (mainly lidar data; sonic data could be involved if available); 4 – Training parameterization models using a machine learning algorithm.

190 2.3 Wind Evolution Model

Following the theoretical considerations by Ropewski et al. (1973), the coherence decreases exponentially with increasing travel evolution time Δt of the signal with respect to “eddy turnover time” τ ÷

$$\gamma_{\text{model}}^2(f) = \exp\left(-C \cdot \frac{\Delta t}{\tau}\right), \quad (2)$$

where the The term C represents the decay behaviour of the coherence depending on the time ratio. C is here virtual which
 195 can could be a constant, a linear function, and even or a more complicated term.

τ is a time scale associated with the characteristic eddy size λ and characteristic velocity of turbulence which is approximated by the standard deviation of wind speed σ as following: follows

$$\tau \sim \frac{\lambda}{\sigma}. \quad (3)$$

This expression implies that eddies are supposed to decay faster under a high-strong turbulence. Given the same degree of
 200 turbulence, large eddies are supposed to take longer time to decay.

The eddy size λ is linked to the frequency of horizontal wind velocity fluctuations f and the flow mean wind speed U with this relation ÷

$$\lambda \sim \frac{U}{f}. \quad (4)$$

Combining Eq. (2)–(4) ~~and introducing the~~, the coherence model becomes

$$\gamma_{\text{model}}^2(f) = \exp\left(-C \cdot \frac{\sigma}{U} \cdot f \cdot \Delta t\right). \quad (5)$$

This equation is essentially the same as the model proposed by Pielke and Panofsky (1970), except that, in their model, Δt is approximated by d/U (d is separation) using Taylor’s (1938) translation hypothesis, indicated as Δt_T .

Simley and Pao (2015a) noted a limitation of this one-parameter model form: the intercept (coherence for 0 frequency) of the modeled coherence is forced to be unity, which is not always realistic. To overcome this issue, Simley and Pao (2015a) introduced a second parameter in the ~~model, as inspired by Simley’s model (2015a), the~~ coherence model, taking a model form similar to the coherence model for transverse and vertical separations suggested by Thresher et al. (1981)

$$\gamma_{\text{model}}^2(f, d) = \exp\left(-a' \sqrt{\left(\frac{fd}{U}\right)^2 + (b'd)^2}\right), \quad (6)$$

where a' and b' are tuning parameters. A comparison between the fitting quality of a one-parameter model and a two-parameter model is given in Sect. 3.2 to confirm the necessity of using a two-parameter wind evolution model.

We have made two modifications to Simley and Pao’s (2015) model. Firstly, d/U is restored to the travel time Δt to avoid coupling Taylor’s (1938) translation hypothesis in the wind evolution model ~~is finally defined as~~–

$$\gamma_{\text{model}}^2 = \exp\left(-\sqrt{a^2 \cdot (f \cdot \Delta t)^2 + b^2}\right),$$

, considering the effect of the wind turbine’s induction zone. In fitting the estimated coherence to the wind evolution model, Δt is determined by the time lag of the peak of the cross-correlation between two wind speed signals, indicated as Δt_M . Secondly, $a'b'd$ is replaced with b . The reasons for that are: 1) With the original form $a'b'd$, $a'b'$ is essentially the fitted term (given that d is known) in the curve fitting. Thus, b' shows a strong dependence on a' , which is generally undesirable for machine learning algorithms. And, 2) the form $a'b'd$ implies that this term is proportional to d , but we found that d is still an important predictor for b' , indicating that the assumption of a linear relationship might be not proper. Therefore, we decided to directly use b to represent the intercept and take d as a predictor instead (see Sect. 2.5).

The modified wind evolution model is

$$\gamma_{\text{model}}^2(f) = \exp\left(-\sqrt{a^2 \cdot (f \cdot \Delta t)^2 + b^2}\right), \quad (7)$$

where the ~~decay parameter a , which summarizes the term C and the other deduced terms,~~ *decay parameter a* represents the decay effect of the coherence, and the ~~offset parameter~~ *offset parameter b* is used to adjust the intercept (coherence for 0 frequency) of the modeled coherence curve. The intercept equals $\exp(-|b|)$. ~~The both wind evolution model~~ Both parameters are dimensionless. The ~~benefit of introducing the~~ *benefit of introducing the* offset parameter b is shown in Sect. 3.2. The travel time Δt is determined by the time lag of the peak of the cross-correlation between two signals, indicated as $\Delta t_{\text{maxcorr}}$. In some previous studies, Δt is approximated by $\frac{d}{U}$ using Taylor’s hypothesis (d is separation), indicated as Δt_{Taylor} . Considering the effect of induction zone

of the wind turbine, this approximation is not applied in this study. The term $f \cdot \Delta t$ is dimensionless, and thus is defined as $f \cdot \Delta t$ is dimensionless, and thus is defined as dimensionless frequency f_{dless} .

235 . In the end, our wind evolution model is defined as

$$\gamma_{\text{model}}^2(f_{\text{dless}}) = \exp\left(-\sqrt{a^2 \cdot f_{\text{dless}}^2 + b^2}\right). \quad (8)$$

In some studies (see e.g. Schlipf et al., 2015), the wind evolution model is defined as a function of wavenumber k , with $k = 2\pi f/U$. The relationship between k and f_{dless} is $k = 2\pi f_{\text{dless}}/d$, applying Taylor's (1938) translation hypothesis. To give an overview intuitive impression of the wind evolution model, Fig. 2 shows the theoretical curves calculated with different

240 values of a and b as example examples.

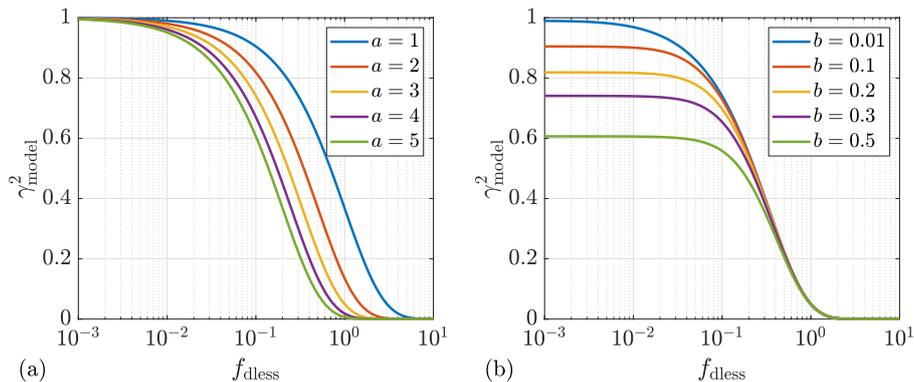


Figure 2. Impact of the model parameters a and b on the wind evolution model. (a) $b = 0$. (b) $a = 3$.

2.4 Estimation of Estimating Coherence using Lidar Data

In this work, the coherence is estimated with lidar data because lidar can provide more data sets with respect to different measuring separations for the estimation of the coherence, which spatial separations. This is not easy to obtain from data of meteorological towers. And when using meteorological towers because multiple towers would be needed and only when the wind direction is aligned with the tower locations would the data be usable. Further, the prediction of the coherence is mainly expected to be applied when coupled with the deployment of a lidar, e.g. in lidar-assisted wind turbine control.

A Doppler wind lidar is a remote sensing device measuring that measures wind speed based on the optical Doppler effect. Lidar emits laser pulses and detects the Doppler shift in backscattered light from the aerosol particles in the atmosphere that are entrained with the wind. The Doppler shift is proportional to the line-of-sight wind speed, i.e. the wind speed projected on onto the laser beam, and thus can be used to estimate the line-of-sight wind speed. The measurement principle of Doppler wind lidar is explained in many publications (see e.g. Weitkamp, 2005; Peña et al., 2013; Liu et al., 2019) (e.g. Weitkamp, 2005; Peña et al., 2013; Liu et al., 2019) and thus is not introduced here in detail.

However, it must be emphasized that the coherence estimated with lidar data ~~is deviated~~ deviates from that estimated with data ~~of point measurements, e.g. from two taken from~~ ultrasonic anemometers. ~~There are several~~ The reasons for that ~~are~~: 1) 255 The sampling rate of ~~a lidar~~ lidars is generally much lower than that of ~~an ultrasonic anemometer~~ ultrasonic anemometers, and ~~thus lidars cannot measure high-frequency fluctuations in wind speed~~; 2) ~~Lidar does not measure the wind speed at a point but the averaged wind speed over the measurement volume along the laser beam~~ the measuring volume of lidars is generally much longer than that of ultrasonic anemometers because of its measurement principle, ~~which is the so-called~~ and thus for lidars, the spatial averaging effect ~~of lidar~~ within the measuring volume needs to be considered; 3) ~~Lidar~~ lidars can only measure the wind 260 speed projected ~~on its~~ onto the emitted laser beams, i.e. the light-of-sight wind speed. The influence of these three aspects is discussed ~~as follows~~ following, specifically considering ~~the case of starring mode of lidar~~ lidar in starring mode:

Low sampling rate of lidar. According to the ~~sampling theorem (Oppenheim et al., 1997)~~ Nyquist–Shannon sampling theorem (Shannon, 1949), the upper frequency limit of a signal transformed from the time domain into the frequency domain is the half of the sampling frequency. As long as the ~~sampling rate of lidar~~ lidar sampling rate is sufficiently high to acquire a complete 265 coherence curve ~~, it will covering the range from the highest coherence (e.g. 0.9–1.0) to the lowest coherence (e.g. 0–0.1), it would probably~~ not have a ~~noticeable effect on the study of the~~ large impact on studying the coherence. To obtain as high a sampling rate as possible, it is decided to select ~~the data of the starring mode~~ starring-mode data to calculate the coherence. ~~The starring mode of lidar generally means~~ Starring mode generally means that the lidar measures the wind speed with a single laser beam pointing ~~towards a fix in a fixed~~ direction. Specifically in this work, the laser beam points ~~to the~~ horizontally upstream of 270 the wind turbine ~~horizontally~~.

Spatial averaging effect of lidar. ~~Consider a pulsed lidar (only pulsed lidars are involved in this work)~~. The spatial averaging effect can be ~~modelled~~ modeled with a moving ~~window averaging average~~ weighted by a Gaussian-like shape function ~~entered at the~~ (see e.g. Carious, 2013) or a triangular function (see e.g. Sathe and Mann, 2012) centered at a measurement point.

~~Consider a pulsed lidar (only pulsed lidars are involved in this work)~~ Following Carious (2013), the weighting function $w(x)$ 275 is an even function centered at every measurement point along the laser beam. The ~~lidar measured~~ lidar-measured wind speed at the measurement point x_0 for any instant can be ~~modelled with~~ modeled with

$$u_1(x_0) = \int_{-\infty}^{\infty} w(x_0 - x)u_p(x)dx = (w * u_p)(x_0), \quad (9)$$

where $u_p(x)$ is a ~~function of the point measurement of the wind speed with respect to the x-axis~~ wind speed function of spatial points on the x-axis aligned with the ~~laser beam of the lidar~~.

280 lidar's laser beam. According to the convolution theorem (~~Oppenheim et al., 1997~~) (?), the following relationship is valid for the Fourier transformation between space and wavenumber domain \div

$$\mathcal{F}\{u_1\} = \mathcal{F}\{w * u_p\} = \mathcal{F}\{w\} \cdot \mathcal{F}\{u_p\}, \quad (10)$$

where $\mathcal{F}\{ \}$ is the Fourier transform operator.

Follow ~~Following~~ Eq. (1), the coherence estimated with lidar data, indicated with the subscript "1", is ~~is~~ is

$$285 \quad \gamma_{ij,1}^2(f) = \frac{|S_{ij,1}(f)|^2}{S_{ii,1}(f) \cdot S_{jj,1}(f)}, \quad (11)$$

where $S_{ii,1}(f)$ and $S_{jj,1}(f)$ are the auto-spectrum at the point i and j , respectively, $S_{ij,1}(f)$ is the cross-spectrum between i and j , and f is the frequency in Hz. They are all estimated ~~with~~ from lidar data. The auto-spectrum is ~~is~~

$$S_{ii,1}(f) = \mathcal{F}\{u_{i,1}(t)\} \cdot \mathcal{F}^*\{u_{i,1}(t)\}, \quad (12)$$

where $u_{i,1}(t)$ is the time series of the wind speed at i , and the symbol $*$ means conjugate. And the cross-spectrum is ~~is~~

$$290 \quad S_{ij,1}(f) = \mathcal{F}\{u_{i,1}(t)\} \cdot \mathcal{F}^*\{u_{j,1}(t)\}. \quad (13)$$

Assume that the laser beam is aligned with the wind direction and ~~the Taylor's hypothesis is valid, (1938) translation~~ hypothesis applies within the measurement volume, and that Eq. (10) is also valid for the Fourier transformation between the time and frequency ~~domain by applying $t = \frac{x}{U}$, U is mean wind speed. Thus, domains.~~ Taylor's (1938) hypothesis is considered valid within the measurement volume because, in principle, wind evolution depends on the evolution time of turbulence (see Eq. (2)), and the measurement volume corresponds to a temporal length on the order of magnitude of 10^{-7} s (typical length of a laser pulse). Now, Eq. (11) can be written as (with t and f ~~are~~ omitted for clarity):

$$295 \quad \begin{aligned} \gamma_{ij,1}^2 &= \frac{|\mathcal{F}\{u_{i,1}\} \cdot \mathcal{F}^*\{u_{j,1}\}|^2}{\mathcal{F}\{u_{i,1}\} \cdot \mathcal{F}^*\{u_{i,1}\} \cdot \mathcal{F}\{u_{j,1}\} \cdot \mathcal{F}^*\{u_{j,1}\}} \\ &= \frac{|\mathcal{F}\{w\} \cdot \mathcal{F}\{u_{i,p}\} \cdot \mathcal{F}^*\{w\} \cdot \mathcal{F}^*\{u_{j,p}\}|^2}{\mathcal{F}\{w\} \cdot \mathcal{F}\{u_{i,p}\} \cdot \mathcal{F}^*\{w\} \cdot \mathcal{F}^*\{u_{i,p}\} \cdot \mathcal{F}\{w\} \cdot \mathcal{F}\{u_{j,p}\} \cdot \mathcal{F}^*\{w\} \cdot \mathcal{F}^*\{u_{j,p}\}}. \end{aligned} \quad (14)$$

Because the function $w(x)$ is real and even, according to the conjugate symmetry of the Fourier transformation (~~Oppenheim et al., 1997~~) (?), $\mathcal{F}\{w\} = \mathcal{F}^*\{w\}$ and $\mathcal{F}\{w\}$ is real and even as well. As a result, all $\mathcal{F}\{w\}$ in the denominator and the numerator are cancelled out. And thus Eq. (14) becomes:

$$300 \quad \gamma_{ij,1}^2 = \frac{|\mathcal{F}\{u_{i,p}\} \cdot \mathcal{F}^*\{u_{j,p}\}|^2}{\mathcal{F}\{u_{i,p}\} \cdot \mathcal{F}^*\{u_{i,p}\} \cdot \mathcal{F}\{u_{j,p}\} \cdot \mathcal{F}^*\{u_{j,p}\}} = \gamma_{ij,p}^2. \quad (15)$$

This means that the spatial averaging effect does not influence the coherence ~~if~~ under the above-mentioned ~~assumptions are fulfilled~~ ideal assumptions.

Misalignment of wind direction and lidar measurement. The above derivation is based on an important assumption that ~~the laser beam is aligned with the wind direction. This will not always be fulfilled in the reality~~ reality, even for a ~~nacelle-based lidar conducting the starring mode.~~

nacelle-mounted lidar operating in starring mode. Figure 3 shows a ~~sketch for a case of misalignment of~~ misalignment between wind direction and lidar measurement direction, at an angle α . The coherence ~~estimated with lidar data of the line-of-sight wind speed~~ is γ_{12}^2 . ~~The coherence of the two longitudinal wind components at the point 1 and 2 and that of~~ the corresponding line-of-sight wind speeds are the same when the both angles are the same, because the angles only introduce

a same constant in the respective wind speed time series. However, γ_{12}^2 is no longer the pure longitudinal coherence only with respect to the longitudinal separation but the horizontal coherence (Panofsky and Mizuno, 1975), combined the effects, which is no longer the longitudinal coherence but the horizontal coherence as defined by Panofsky and Mizuno (1975). γ_{13}^2 and γ_{23}^2 are the longitudinal and lateral coherence, respectively.

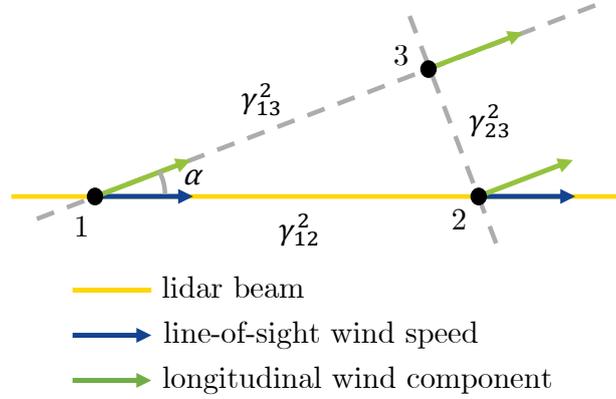


Figure 3. Misalignment of wind direction and lidar measurement. α is the misalignment angle. γ_{12}^2 is the coherence of the line-of-sight wind speed. γ_{13}^2 and γ_{23}^2 are the longitudinal and lateral coherence, respectively.

315 Schlipf et al. (2015) suggested a model for the horizontal coherence (magnitude coherence) based on the assumption of point-measurement for simplification

$$\gamma_{ij, \text{loSP}} = \frac{\cos^2(\alpha) \gamma_{ij, \text{ux}} \gamma_{ij, \text{uy}} S_{ii, \text{u}}}{\cos^2(\alpha) S_{ii, \text{u}} + \sin^2(\alpha) S_{ii, \text{v}}}, \quad (16)$$

where $\gamma_{ij, \text{loSP}}$ is the horizontal coherence of line-of-sight wind speed point-measurements, $\gamma_{ij, \text{ux}}$ and $\gamma_{ij, \text{uy}}$ are the longitudinal and lateral coherence of the longitudinal wind component, $S_{ii, \text{u}}$ and $S_{ii, \text{v}}$ are the auto-spectra of the longitudinal coherence γ_{13}^2 and the lateral coherence γ_{23}^2 . To retrieve and lateral wind components. Based on this equation, determining the longitudinal coherence in this case, the $\gamma_{ij, \text{ux}}$ is possible only given a specific turbulence model (knowing $S_{ii, \text{u}}$, $S_{ii, \text{v}}$ and $\gamma_{ij, \text{uy}}$) and knowing the misalignment angle α . Moreover, the above discussed spatial averaging effect must be coupled to a specific turbulence model (Schlipf, 2015; Mann et al., 2009), and thus the wind evolution model is the horizontal coherence, considering that the lateral coherence for the point at x depends on the lateral separation Δy associated with its distance from the center point of the range gate x_0 , i.e. $\Delta y = \cos(\alpha)(|x - x_0|)$. Therefore, the longitudinal coherence is implicitly included in the final model implicitly integration of horizontal coherence weighted by the range weighting function of lidars.

Sketch of misalignment of wind direction and lidar measurement:

In this study, it is decided to develop the we decide to develop a parameterization model based on the horizontal coherence for the following reasons: First, a. Firstly, consider the case for a nacelle-mounted lidar. The misalignment of the lidar measurement means that the wind turbine is misaligned as well. In this case, it makes sense to predict the corresponding

horizontal coherence. Secondly, a standalone parameterization model, independent from any turbulence model, is desired for more flexibility in application. ~~Second, the determination of the model parameters of~~ Thirdly, determining the parameters in an implicit wind evolution model is complicated ~~and it is a must~~ when using measured data. And it is necessary to acquire the ~~angle between the wind direction and the laser beam~~ misalignment angle α , which is not always possible in application, especially when ~~considering lidar as~~ lidar is the only data source~~-, though deployment of lidars with multiple beams might help in this case~~. Moreover, the requirement for the accuracy of α is very high because α is included in the most basic step — fitting the estimated coherence to the wind evolution model. The uncertainties contained in α will propagate through the whole model and affect the further analysis radically. Since the prediction concept ~~is expected~~ needs to be applicable ~~for different data availability~~. ~~It under different data availabilities, it~~ is not desired to make the ~~estimation of coherence depend~~ fitting process depend so critically on a variable ~~which is whose~~ availability and accuracy are not always guaranteed ~~to be available~~. Third, consider the case for a nacelle-based lidar, the presence of the misalignment of the lidar measurement means the wind turbine misalignment exists as well. In this case, it makes sense to predict the corresponding horizontal coherence. ~~Whereas, the variation-~~ It is thus helpful to consider α as a predictor (see Sect. 2.5) to account for variations in the horizontal coherence caused by the direction misalignment ~~is attributed to the misalignment angle, which is considered as one of the predictors~~. The benefit of doing so is to make α more standalone and to prevent its errors from affecting everything else, while reasonably taking its influences into account. In addition, Gaussian process regression inherently assumes imperfect training data (containing noisy terms; see Sect. 2.5)2.6), so it is better to keep uncertainties in predictors.

Certainly, if the ~~measurement of the~~ direction misalignment is available and sufficiently accurate in a given application scenario, the prediction concept can be easily adjusted by changing the ~~model used in the second step in the workflow (see Sect. 2.2) to obtain the~~ wind evolution model parameters to which the estimated coherence is supposed to fit.

2.5 Potential Predictors

~~Three groups of wind field condition variables are supposed to be~~ In the literature reviewed in the introduction, the variables considered relevant to wind evolution ~~-, and thus will be the~~ are as listed below:

- Ropelewski et al. (1973): turbulence intensity (a function of roughness length and Richardson number (J. L. Lumley and H. A. Panofsky))
- Panofsky and Mizuno (1975): mean wind speed, turbulence intensity, standard deviation of the lateral wind component, lateral integral length scale of the longitudinal wind component, longitudinal separation, and the angle between the wind direction and the measurement line (if misalignment exists)
- Kristensen (1979): turbulence intensity, longitudinal integral length scale of the longitudinal wind component, and longitudinal separation
- Simley and Pao (2015a): turbulence intensity, longitudinal integral length scale of the longitudinal wind component, and longitudinal separation

The above-mentioned variables can be categorized into three groups: wind statistics, atmospheric stability, and relative positions of measurement points. We follow this train of thought to discuss potential predictors of the surrogate model. A discussion about these variables is presented as follows: parameterization models. It is worth mentioning, in advance, that not all of these predictors will be used in the final models. Useful features will be selected using the automatic relevance determination squared exponential kernel function (Duvenaud, 2014). The goal of this initial step is to collect all possible predictors even though some of them will turn out to be redundant and can be converted to each other.

Wind statistics. As explained in Sect. 2.3, wind evolution is supposed to be correlated with the degree of turbulence or rather turbulent kinetic energy. Therefore, Following prior research, turbulence intensity I_T is considered as a predictor. The turbulence intensity is defined as

$$I_T = \frac{\sigma}{U}. \quad (17)$$

In addition, mean wind speed U and its standard deviation σ , turbulence intensity are also included because they are the fundamental variables of turbulence intensity. Apparently, I_T and σ are equivalent (given U), so only one of them will be selected according to the result of feature selection.

Moreover, integral length scale L and integral time scale T are considered as representative variables. The integral time scale T is defined as (Pope, 2000): is considered as a predictor, and approximated with (Pope, 2000; Simley and Pao, 2015a)

$$TL = U \cdot \int_0^{\infty} \rho(s) ds = U \cdot T, \quad (18)$$

where $\rho(s)$ is the autocorrelation function. In this study the Indeed, integrating the autocorrelation gives the integral time scale is computed by integrating the autocorrelation T , and thus T is also considered as a predictor. L and T thus constitute another pair of redundant predictors from which only one will be selected. The integration of autocorrelation is computed up to the first zero crossing. Since it is not possible to calculate the integral length scale zero-crossing instead of infinity in practice (Simley and Pao, 2015a). It is worth mentioning that the approximation of L based on its definition, it is approximated with the integral time scale T :

$$L = U \cdot T.$$

essentially applies Taylor's (1938) translation hypothesis, and thus L might contain uncertainties.

Moreover Besides the variables already considered in prior studies, it is interesting to study the influence of other statistical characteristics on wind evolution, such as explore whether high-order wind statistics such as skewness and kurtosis of wind speed could play a role in wind evolution prediction. Skewness (i.e. the third standardized central moment skewness $\tilde{\mu}_3$, a measure) and kurtosis (i.e. the fourth standardized central moment) are measures of the asymmetry of the distribution of the wind speeds in a data block, and the fourth standardized central moment kurtosis $\tilde{\mu}_4$, a measure of the "tailedness", and flatness of the wind speed distribution, respectively. The sample skewness G_1 , with bias correction, is defined as (Joanes and Gill, 1998)

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \cdot \frac{\frac{1}{n} \sum_{i=1}^n u_i^3}{\left(\frac{1}{n} \sum_{i=1}^n u_i^2\right)^{3/2}}, \quad (19)$$

395 and the sample kurtosis G_2 (not subtracting 3), with bias correction, is defined as (Joanes and Gill, 1998)

$$G_2 = \frac{n-1}{(n-2)(n-3)} \cdot \left[(n+1) \cdot \frac{\frac{1}{n} \sum_{i=1}^n u_i^4}{\left(\frac{1}{n} \sum_{i=1}^n u_i^2\right)^2} - 3(n-1) \right] + 3, \quad (20)$$

where u_i is wind speed fluctuations, and n is the number of data points. The sample skewness and kurtosis determined from measured data would probably contain large uncertainties. For example, Lenschow et al. (1994) found that statistical moments estimated using time series data with limited length show a systematic deviation from the true moments. Despite this, it is still
400 worth investigating whether these two high-order wind statistics could be useful for prediction.

Atmospheric stability. The atmospheric stability represents a global effect of the surface layer in the boundary layer on the wind field, and thus it is considered as a wind field. It is believed to affect wind evolution being an influence factor on the wind evolution. The atmospheric stability is usually characterized with Monin-Obukhov length or Richardson number. In this study a turbulence stability (Ropelewski et al., 1973; J. L. Lumley and H. A. Panofsky, 1964). A dimensionless height ζ built with Monin-Obukhov, built with Obukhov length L_{MO} is used as stability indicator. Because it is found a functional relationship between the dimensionless height ζ and the gradient Richardson number for flat terrain (Businger et al., 1971), these two variables are supposed to be equivalent for representing the influence of stability in the surrogate model. The dimensionless height ζ is defined as (Businger et al., 1971): (Obukhov, 1971), is considered as a predictor (Businger et al., 1971).
405

$$\zeta = \frac{z}{L_{MO}} = -\frac{\kappa g \overline{w'\theta'_v} z}{\overline{\theta u_*^3}}, \quad (21)$$

410 where κ is the von Kármán constant, g is gravitational acceleration, z is the measurement height, $\overline{\theta}$ is average of the mean potential temperature, u_* is the friction velocity, and $\overline{w'\theta'_v}$ is covariance of perturbations of vertical velocity the covariance of vertical velocity perturbations and virtual potential temperature.

Relative position of the measurement points. Relative positions of measurement points. The distance between the two measurement points Based on our modifications to Simley and Pao's (2015a) model (see Sect. 2.3), measurement separation d is likely to play an important role in has been removed from the wind evolution because it determines how far the eddies travel, and thus how likely or to what extent the local terrain changes. So as the travel time Δt . For prediction, it is not possible to obtain $\Delta t_{\max\text{corr}}$. Therefore, the travel time approximated using Taylor's Hypothesis Δt_{Taylor} is considered as predictor.
415

model and is now considered as a predictor. As discussed in Sect. 2.4, the angle between the lidar beam and the main wind direction α is associated to misalignment angle α is not involved in fitting the wind evolution model but is considered as a predictor to account for the influence of the lateral coherence on the horizontal coherence. In fact, d is associated with two different effects. On the one hand, d corresponds to travel time or, rather, to evolution time Δt , and thus is considered as a relevant variable. And it is important to analyze to what extent the horizontal coherence would depend on this angle.
420

It is emphasized that the variables discussed above are "potential" predictors of the parameterization model according to physical consideration. Further analysis needs to be done to determine which of them should be selected for model training, which is called *feature selection*. The principle of feature selection is to figure out variables which provide best predictive power (accounting to most of the variation of the model responses) and ideally, these variables should be independent from each other to prevent over-fitting in model training. Since one of the objectives is to figure out the necessary predictors under different data availability, different combinations of predictors are discussed in Sect. 5, believed to play an important role in wind evolution. On the other hand, d together with α account for the decay of the lateral coherence. The travel time determined with the maximum cross-correlation Δt_M is a more accurate variable. However, considering that calculating Δt_M might not always be feasible due to its computational complexity, the travel time approximated using Taylor's (1938) translation hypothesis Δt_T is included as well.

The Table 1 lists the notations of the potential predictors used in this work. Above-mentioned potential predictors are summarized in Table 1. These variables are acquired from data of lidar and/or ultrasonic anemometers according to the data availability of the derived from both lidar data and data measured with ultrasonic anemometers (hereafter referred to as sonic data) according to their availability in each measurement campaign. This is distinguished with subscripts. The measurement instrument is indicated with a subscript: "l" for lidar and "s" for ultrasonic anemometer (i.e. ultrasonic anemometer). For example, U_l means represents the mean wind speed calculated with lidar data while U_s means that calculated with data from ultrasonic anemometer (hereinafter for brevity, also referred to as sonic data). Regarding sonic data, it is more reasonable for the analysis of wind evolution to use a wind coordinate system with the x -axis aligned to the mean wind direction instead of the meteorological coordinate system. The mean wind direction is determined with the mean wind direction for each data block. The high-resolution longitudinal (indicated with the subscript "x") and lateral (indicated with the subscript "y") wind speeds are obtained by projecting the high-resolution wind components measured with ultrasonic anemometers on the wind coordinate system. Then, the above-mentioned variables are derived from the data based on the wind coordinate system. For example, $U_{x,s}$ represents the mean wind speed calculated from the longitudinal wind component measured with ultrasonic anemometers.

2.6 Gaussian Process Regression

In the preliminary study it is found out that the Gaussian Process Regression (GPR) overall performs the best for the prediction of the wind evolution model parameters (Chen, 2019). Therefore, in the present work, attempt is made to further analyze the applicability of the GPR model for this purpose taking into account different measurement scenarios. This section briefly introduces the principle of the Gaussian process regression (GPR) and the most important model hyperparameters which modify the behaviour hyperparameters that modify the behavior of a GPR model. The model training is done using Matlab the MATLAB Statistics and Machine Learning Toolbox¹.

The principle of GPR. Think of Consider making a regression model from some data. A very intuitive approach is fitting the data to certain types of function intuitive approach is to fit certain functions, e.g. linear function or polynomial function or polynomial. However, this requires an initial guess about the underlying functions of functional relationship(s) behind the

¹<https://de.mathworks.com/products/statistics.html>

Table 1. Notations of potential predictors.

<u>Notation</u>	<u>Variable</u>	<u>Unit</u>
U	mean wind speed	$[\text{ms}^{-1}]$
σ	standard deviation of wind speed	$[\text{ms}^{-1}]$
$\tilde{\mu}_3 G_1$	skewness of wind speed	[-]
$\tilde{\mu}_4 G_2$	kurtosis of wind speed	[-]
I_T	turbulence intensity	[-]
T	integral time scale	[s]
L	integral length scale	[m]
ζ	dimensionless Obukhov length	[-]
d	distance-measurement separation	[m]
α	angle between lidar-beam and main-wind-direction <u>wind direction and lidar measurement</u>	[°]
Δt_M	<u>travel time — maximum cross-correlation</u>	[s]
Δt_T	<u>travel time — Taylor's (1938) approximation</u>	[s]

data, which is very difficult in this case ~~because~~ because the wind evolution model parameters do not indicate any clear ~~dependency~~ dependence on the potential predictors. The reasons for that could be multiple: ~~first, 1)~~ the data could be noisy; ~~second, the dependency and 2) the dependence~~ could exist in multidimensional space ~~which is not possible to be observed~~ not observable in a single dimension, etc. Under this circumstance, GPR turns out to be a good choice because it is non-parametric probabilistic model, which means the model is not a ~~"fix"-function~~ specific function, but a probability distribution over functions. The principle ~~of the GPR is based on the underlying GPR is~~ Bayesian inference. The prior distribution over functions, which can be understood as a guess about what ~~kind of functions~~ kinds of function could be present without knowing the data, is specified by a particular Gaussian process (GP) which ~~favours~~ favours smooth functions. In the training process, as adding the data, the ~~distribution over functions is adjusted in the way that the probability of the~~ probabilities associated with the functions which do not agree with the observations will be decreased, which gives the posterior distribution over the functions (Rasmussen and Williams, 2006).

Hyperparameters of GPR. The ~~behaviour~~ behavior of a GPR model is defined by its hyperparameters. To introduce the ~~hyperparamters~~ hyperparameters, a basic explanation ~~in mathematical aspect~~ is given following Rasmussen and Williams (2006). Please note that the complete deduction is not displayed here because it is beyond the scope of this paper. For further details, please refer to Chapter 2 of Rasmussen and Williams' (2006) book.

~~The~~ GPR is based on ~~the~~ Bayesian inference. First, consider a single observation. The Bayesian linear regression model with Gaussian noise is defined as:

$$f(\mathbf{x}) = \phi(\mathbf{x})^\top \mathbf{w}, \quad y = f(\mathbf{x}) + \varepsilon, \quad (22)$$

where \mathbf{x} is ~~the input vector of different parameters~~ an input vector containing D different predictors of a single observation,
 475 $\phi(\mathbf{x})$ is the function which maps the input vector ~~into~~ onto a higher dimensional space where the Bayesian linear model is applicable, \mathbf{w} is ~~the weight vector of~~ a vector of weights of the linear model, $f(\mathbf{x})$ is the function value, y is the observed target value, and ε is independent identically distributed Gaussian noise with zero mean and variance σ_n^2 :

$$\varepsilon \sim \mathcal{N}(0, \sigma_n^2). \quad (23)$$

The Bayesian linear model is a GP given ~~the prior that the prior distribution~~ of \mathbf{w} is Gaussian normally distributed with zero
 480 mean. Since a GP is fully specified by its mean and covariance, ~~it~~ the Bayesian linear model is written as \div

$$f(\mathbf{X}) \sim \mathcal{GP}(\mathbf{0}, \text{cov}(f(\mathbf{X}))), \quad (24)$$

where \mathbf{X} is the aggregation of all input vectors of n observations. This is the prior distribution over functions. The presence of ε shows another advantage of ~~the GPR~~ GPR, viz. that it is able to inherently assume noisy observations and take ~~into account~~ this effect this effect into account in the model. σ_n is one of the hyperparameters.

485 It is common, but not necessary, to assume GPs with a zero mean function. The ~~use~~ mean function can be modeled with a set of basis functions ~~makes it possible to specify a non-zero mean over functions~~ $\mathbf{h}(\mathbf{x})$ and a corresponding coefficient vector β . So, ~~the model can also~~ GPs with a non-zero mean function can be assumed as \div

$$g(\mathbf{x}) = f(\mathbf{x}) + \mathbf{h}(\mathbf{x})^\top \beta, \quad (25)$$

~~where $\mathbf{h}(\mathbf{x})$ are a set of basis functions and are the corresponding coefficients.~~ Basis The basis function is one of the hy-
 490 perparameters. ~~The coefficients are estimated with~~ MATLAB provides four types of basis function: zero (assuming no basis function), constant, linear, and pure quadratic. The coefficient vector β is estimated from training data.

The covariance of the function values is ~~usually acquired through~~ not specified explicitly but estimated using a kernel function \div

$$\text{cov}(f(\mathbf{X})) = \text{K}(\mathbf{X}, \mathbf{X}), \quad (26)$$

495 which is the so-called kernel trick. There are two types of kernel functions: one is kernel functions with the same characteristic length scale for ~~each predictor~~ all predictors; the other ~~is that with~~ has separate characteristic length ~~scales~~ scales. The latter ~~is called Automatic Relevance Determination kernel functions which are called automatic relevance determination kernel functions and~~ can be used to select predictors. ~~Kernel~~ The kernel function and its characteristic length scale(s) are hyperparameters of the GPR model.

500 In this work, ~~the~~ Automatic Relevance Determination Squared Exponential kernel function (ARD-SE kernel) (Duvenaud, 2014) is applied. The ARD-SE kernel function is basically ~~the~~ a squared exponential kernel function (SE kernel) with a separate characteristic length scale σ_m for each predictor m (m is the index of predictors). For any pairs of observations i, j , the ARD-SE kernel function is defined as \div

$$\text{K}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_j^2 \exp \left[-\frac{1}{2} \sum_{m=1}^D \frac{(x_{im} - x_{jm})^2}{\sigma_m^2} \right], \quad (27)$$

505 where σ_f^2 denotes the signal variance, which determines the variation of function values from their mean. The characteristic length scale σ_m implies the sensitivity of the function being modeled to the predictor m . A relatively large length-scale σ_m indicates a relatively small variation along the corresponding dimensions in the function, which means these predictors are less relevant in-comparison-to than the others (Duvenaud, 2014).

In the end, the key predictive equation for GPR can be derived by conditioning the joint Gaussian prior distribution on the
510 observations ;, and it is normally distributed.

$$\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*)), \quad (28)$$

where the subscription * denotes test data \mathbf{X}_* denotes new input data used in prediction. \mathbf{f}_* represents $\mathbf{f}(\mathbf{X}_*)$ for convenience, which is the predicted function value.

To summarize, the hyperparameters defining the a GPR model are the basis function $\mathbf{h}(\mathbf{x})$, the noise standard deviation
515 of the Gaussian process model σ_n , the kernel function $K(x_i, x_j)$, the standard deviation of the function values σ_f , and the characteristic length scale in the kernel function σ_m . These hyperparameters can be tuned-tuned in the training process to achieve a better model.

2.7 Model Validation

The trained model is evaluated with a k-fold k-fold cross-validation ; in which the data is divided into k-k disjoint, equally
520 sized subsets. The model validation is done with one subset (also called in-fold observations) and the training is done with the rest k-1 remaining $(k-1)$ subsets (also called out-of-fold observations). This procedure is repeated k-k times, each time with a different subset for validation. The predicted target values and the goodness-of-fit measures of the regression models are computed for in-fold observations using a model trained on out-of-fold observations.

Theoretically, k can be any integer between two and the number of observations (a special case called 'leave-one-out').
525 When k is very small, the sample size of training data ($\frac{k-1}{k}$ of the total observations) could be insufficiently large. However, considering that the training process must be repeated k times, it would take a very long time when k is very large. As a compromise between these two factors, k is commonly set to 5-10 in machine learning. In this study, 5-fold cross-validation was is applied.

The model performance is evaluated with two goodness-of-fit measures: the Root-Mean-Square-Error root-mean-square error
530 (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i^N (y_i - y_{\text{pred},i})^2} \quad (29)$$

and the coefficient of determination (R^2) ; which are defined with Eq. (29) and Eq. (30), respectively.

$$\text{RMSE } R^2 = \sqrt{\frac{1}{N} \sum_i^N (y_i - y_{\text{pred},i})^2} \frac{1}{\sqrt{1 - \frac{\sum_i^N (y_i - y_{\text{pred},i})^2}{\sum_i^N (y_i - \bar{y})^2}}}, \quad (30)$$

535
$$R^2 = 1 - \frac{\sum_i^N (y_i - y_{pred,i})^2}{\sum_i (y_i - \bar{y})^2},$$

In Eq. (29) and Eq. (30) where y and y_{pred} denote the observed target values and predicted target values respectively, \bar{y} denotes the average of the observed target values, N denotes the number of observations.

It is worth mentioning that, according to this definition, R^2 can be understood as taking prediction with the mean value of the observations as reference a reference by which to evaluate the model performance. In this case, R^2 ranges from $-\infty$ to one, for perfect prediction. R^2 yields equals zero if the prediction is simply made with the mean value of the observations. The higher the R^2 is, the better the model performs. A negative value of R^2 indicates that the prediction with the selected model is selected model performs even worse than the prediction just using prediction using just the mean value of the observations.

540

3 Data Processing

545 This section first introduces the data sources in Sect. 3.1 and then explains the procedure for the determination of the wind evolution model parameters in Sect. 3.2.

3.1 Data Source

In this study, This study involves measured data from two research projects are involved. The reasons for using two different data sources are, on the one hand, to find commonality in-between two different measurements and avoid accidental conclusions, and, on the other hand, to study whether there are differences or what kind of differences in the wind evolution can be observed. The relevant research projects as well as the measurement campaigns are briefly introduced (briefly) as follows:

550

LidarComplex. The research project LidarComplex was funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). In this project, a lidar measurement campaign was carried out in Grevesmühlen, Germany. The measurement site is basically flat, mainly farmland with hedges and few large trees. The More details about the measurement campaign can be found in Schlipf et al. (2015). The lidar deployed in this measurement campaign was the SWE Scanner 1.0, which was adapted from a WindCube V1 from Leosphere (Schlipf et al., 2015). This lidar has five measurement range gates (can be understood as measurement distances), focusing at distances of 54.5 m, 81.75 m, 109 m, 136.25 m, and 163.5 m, respectively. The full width at half maximum (FWHM) of the measurement range gates is 30 m (Carious, 2013). The lidar was installed on the nacelle of a wind turbine with a rotor diameter of 109 m in the measurement campaign. A (rotor diameter of 109 m) at 95 m. In addition, a meteorological mast is located 295 m southwest of the wind turbine on which the lidar was installed. The data from an ultrasonic anemometer installed at 93 m on the meteorological mast is also involved in this study. SCADA data of the wind turbine is also available. More details about the measurement campaign can be found in Schlipf et al. (2015).

560

The data of the ultrasonic anemometer installed on 93 m height is used in this study. Instead of the meteorological coordinate system, the wind coordinate system with x-axis aligning to the main wind direction is more useful for the analysis of wind

565 evolution. Thus, the longitudinal (indicated with the subscript "x") and the lateral (indicated with the subscript "y") wind speeds are obtained by projecting the high-resolution time series wind speed on this coordinate system. The main wind direction is determined with the Recorded yaw positions are used to estimate the misalignment angle α , assuming that the mean wind direction at the turbine can be approximated with the mean wind direction for each data block measured on the meteorological mast.

570 **ParkCast.** The ~~research project~~ ParkCast² project is an ongoing project funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). ~~Currently~~ While this paper is in preparation, a lidar measurement campaign is being conducted on the offshore wind farm ~~alpha-ventus~~ *alpha-ventus*³. Two long-range lidars ~~StreamlineXR~~ (StreamlineXR) have been deployed in the measurement campaign. The data used ~~in this study was gathered by the one~~ here is from the lidar installed on the nacelle of ~~the~~ wind turbine AV4 (rotor diameter of 126 m) at 92 m, measuring the inflow ~~stream~~. Unfortunately, the data of the ~~the~~ measurement distances were set to 30 m to 990 m with an increment of 60 m. The FWHM of the measurement range gates is 60 m. Unfortunately, neither data from the meteorological mast on FINO1 ~~and the~~⁴ nor SCADA data of ~~the~~ AV4 for the observed period was ~~not yet~~ available when the analysis was done. Therefore, ~~it has to be assumed that the wind turbine had no direction misalignment during the observed period and the lidar was always measuring along the main wind direction~~ the misalignment angle α is not available for ParkCast.

580 ~~In comparison~~ Compared to ultrasonic anemometers, lidar systems have much lower sampling ~~rate because of its measurement principles~~ rates. To obtain ~~as high sampling rates as possible, it is decided to~~ the highest possible sampling rate, we select the measurement periods where the staring mode was ~~conducted for the used, for~~ both campaigns.

Essential information about the measurements ~~are summarized in the~~ is summarized in Table 2. Figure A1 gives an overview of the wind statistics of these two selected measurement periods by illustrating the relative frequency distribution of ~~lidar measured~~ lidar-measured wind speed and ~~lidar measured~~ turbulence intensity. For brevity, "~~LidarComplex~~" and "~~ParkCast~~" 'LidarComplex' and 'ParkCast' are used to refer to the selected measurements throughout the paper.

3.2 Determination of Wind Evolution Model Parameters

To obtain the wind evolution model parameters a and b , the wind evolution is estimated with lidar data and then fitted to the wind evolution model (Eq. 58). The processing procedure is described as follows:

590 **Step 1: Filtering of the lidar data.** *Step 1: Filtering of the lidar data.*

The lidar data from LidarComplex is filtered ~~using a CNR filter with the valid range from -24 dB to -5 dB~~ according to the carrier-to-noise ratio (CNR) of the lidar signals (*CNR filter*). The valid range of the CNR filter is -24 dB to -5 dB, determined from the plot of CNR values and wind speed.

595 ~~The lidar data from ParkCast is filtered with a range filter which is originally applied in image processing but adapted~~ A CNR filter is not, however, suitable for lidar data ~~filtering for the~~ from ParkCast because, for a long-range lidar by Würth et al. (2018)

²<https://www.rave-offshore.de/en/parkcast.html>

³<https://www.alpha-ventus.de/english>

⁴<https://www.fino1.de/en/>

Table 2. Summary of measurement setups.

<u>Measurement campaign</u>	LidarComplex	ParkCast
Selected period	02 Dec 2013–20 <u>2013–20</u> Dec 2013	04 June–Jun <u>Jun 2019–14 Jun</u> 2019 –14 June 2019–
Location	Grevesmühlen, Germany	Alpha-Ventus <u>alpha ventus</u>
Terrain type	onshore, flat	offshore
Device	nacelle based lidar + met mast	nacelle based lidar
Measurement height [m]	95 (lidar), 93 (sonic)	92
Range gate [m]	54.5; 27.25; <u>81.75;</u> ... 163.5	30; 60; <u>90;</u> ... 990
Number of range gates	5	17
<u>Full width at half maximum [m]</u>	<u>30</u>	<u>60</u>
Sampling rate [Hz]	0.99	0.27
Valid samples* <u>Valid samples*</u>	3285	10112

* After lidar data filtering, data pairing, and outlier filtering. For details see Sect.3.2

~~In comparison to the CNR filter, the range filter removes less valid data points while ensuring the filter quality. Its approach is to detect the extreme values which exceed the threshold within~~, the backscattered signals from distant range gates could be very weak, and thus the CNR values could be low even when the measured wind speed is plausible. Würth et al. (2018) suggested an approach to filter the data based on the value range (*range filter*) and the standard deviation (*standard deviation filter*) within a certain number of adjacent data points defined ~~by the window size~~ as a window, which can keep more valid data than a CNR filter. A range filter ~~is~~ detects the maximum value difference within a window and filters the data points for which the maximum value difference exceeds a threshold. A standard deviation filter calculates the standard deviation within a window and filters the data points for which the standard deviation exceeds a threshold. Both filters are applied to check the line-of-sight wind speed ~~of lidar and its standard deviation. The thresholds for both are 6 m/s and 3 m/s~~ with thresholds of ~~6 ms⁻¹ and 3 ms⁻¹~~, respectively. The window size is set ~~as to~~ three data points.

~~Step 2: Estimation of coherence.~~ Step 2: Estimation of coherence.

The lidar data is divided into ~~30-min~~ 30-minute blocks. This is ~~also~~ consistent with the ~~common~~ commonly used period for ~~ealeulation of~~ calculating the Obukhov length. Only the data blocks with more than 80% valid data points are used to estimate the coherence. The missing values are ~~interpolated with the~~ estimated by shape-preserving piecewise cubic interpolation ~~And~~ ~~the~~ (Fritsch and Carlson, 1980). The missing end values are each replaced with their nearest value.

The ~~data~~ Data measured at different range gates (i.e. measurement ~~distance~~ distances) is paired in the way shown in Fig. 4 to obtain as many samples (i.e. data blocks) as possible. The pairing has ~~C_N^2 possibilities~~ ~~($N - \binom{N}{2}$) possibilities~~ (N is the number of the lidar range gates).

The travel time of the wind field is approximated with the time lag at the maximum of the cross-correlation $\Delta t_{\max\text{corr}}$ between these two wind speed signals. The upstream point is always regarded as the reference point. The data measured at the downstream point is shifted by $\Delta t_{\max\text{corr}}$

Δt_M to match the reference wind speed data. The magnitude-squared coherence is estimated using Welch's overlapped averaged periodogram method using a Hamming window, 24 segments, and 50% overlap.

The data of the reference point is used to calculate the lidar-measured lidar-measured wind statistics.

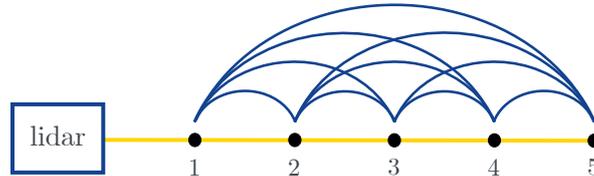


Figure 4. Pairing of different measurement points for estimation of coherence for LidarComplex, given as an example.

620 **Step 3: Fitting to the wind evolution model.** *Step 3: Fitting to the wind evolution model.*

Because Before fitting the model, we must consider two issues that might introduce noise into the coherence estimate. Firstly, because both lidars are installed on the top of the wind turbine nacelle and the nacelle is actually moving with certain frequencies nacelle of a wind turbine which is actually in motion, the focus points of lidar are thus the laser beams are moving as well. This movement motion causes excitation at certain frequencies in the estimated coherence.

625 Figure A2 shows a comparison between an example coherence curve and the power spectral density (PSD) of the fore-aft and in-plane tower top acceleration of LidarComplex. The excitation in the coherence conforms to that in the both PSDs and is mainly located in the frequency range higher than occurs mainly at frequencies above 0.2 Hz. To avoid the negative effect negative effects on the fitting quality caused by this excitation, the cut-off frequency is set to be 0.2 Hz hence set at 0.2 Hz, and the coherence is only fitted fitted only up to this cut-off frequency.

630 Secondly, according to Schlipf (2015), critical wavenumbers where the lidar signals would be only determined by noises must be checked. The critical wavenumbers are $2\pi/W_L$ (W_L is the full width at half maximum of the range gate) and its harmonics. As mentioned in Sect. 2.3, the relationship between wavenumber k and dimensionless frequency f_{dless} is $f_{\text{dless}} = kd/2\pi$. Thus, the smallest critical value of f_{dless} is d/W_L . Considering LidarComplex as an example, $W_L = 30$ m and $d = 27.25$ m for the smallest separation, which is the most critical case. $d/W_L \approx 0.91$, which is already located in the filtered part (see the grey area in Fig. 5 (a)).

The fitting is done through nonlinear least squares fitting using by a nonlinear least-squares method using the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963; Moré, 1978). Only the data blocks with R^2 of the fitting higher than 0.8 $R^2 > 0.8$ are considered as valid samples.

Step 4: Outliers filtering. *Step 4: Outlier filtering.*

640 The final filtering was done by checking the value distribution of every parameter-relevant variable to omit outliers. It is emphasized that outlier is not necessarily wrong data. For outliers are not necessarily false data. In some cases, outlier is just sample collected from the value range where the outlier is from a value range in which not enough samples have been collected. Though, it were collected. It is very important to filter the outliers properly because it is difficult for the a regression model to capture the relationship for those value ranges with too few samples. The outliers are determined as the data with value higher
645 than Because the distributions of the variables all have a long right tail, the outliers are chosen as all data exceeding the 99th percentile of the data.

Figure 5 is an example plot of the data block from 07 Dec. 2013, 12:00-12:00-12:30, from LidarComplex. This data block is selected here for two reasons: data integrity and representative wind statistics. In this data block, the lidar-measured mean wind speed is 7.3 ms^{-1} to 7.7 ms^{-1} , and the lidar-measured turbulence intensity is 0.10 to 0.12, for different range gates. These
650 values appeared frequently in the selected period according to Fig. A1. Hence, this data block is regarded as a representative ease study case study example for LidarComplex and is referred throughout to throughout the paper. The figure illustrates the estimated coherence between different range gates and the corresponding fitted curves. The shaded areas show the that the selected cut-off frequency of 0.2 Hz is reasonable for this case. A similar plot from ParkCast is found in Fig. A3. Because the sampling rate of ParkCast is lower, the excitation by the nacelle top's movement is not observed in the coherence, and thus no
655 cut-off frequency was set for ParkCast data.

In Fig. 5 (c) and (d), the intercept of the coherence is much lower than one-1 even though the separation is not very large. This confirms the necessity of choosing a wind evolution model which is able to define different offset values depending on the conditions. Indeed, in comparison to compared with the fitting quality of Pielke and Panofsky's model which merely contains contains merely a single parameter — the decay parameter a ; — the fitting quality of the wind evolution model (Eq. (58)) is
660 overall better (see Fig. A4). The R^2 of the fitting to value of R^2 for the fitting of Eq. (58) is almost always higher than that of the fitting to the for the fitting of Pielke and Panofsky's (1970) model. The wind evolution model used in this work (Eq. (58)) is thus proven to be able to model the coherence better.

4 Statistical Analysis of Wind Evolution

This section presents the results of the a statistical analysis of the wind evolution, including the distributions of the wind
665 evolution model parameters ; in (Sect. 4.1 and their dependency on the measuring separation in) and their dependence on measurement separation (Sect. 4.2).

4.1 Distribution of the Wind Evolution Model Parameters

To study the overall characteristics of the wind evolution, the value distributions of the wind evolution model parameters for the both measurements are displayed in Fig.6.

670 As listed in Table 2, there are two main differences between the lidar settings in the both measurements: sampling rate and measurement range, which might affect the distributions of the wind evolution parameters. To enhance the comparability of

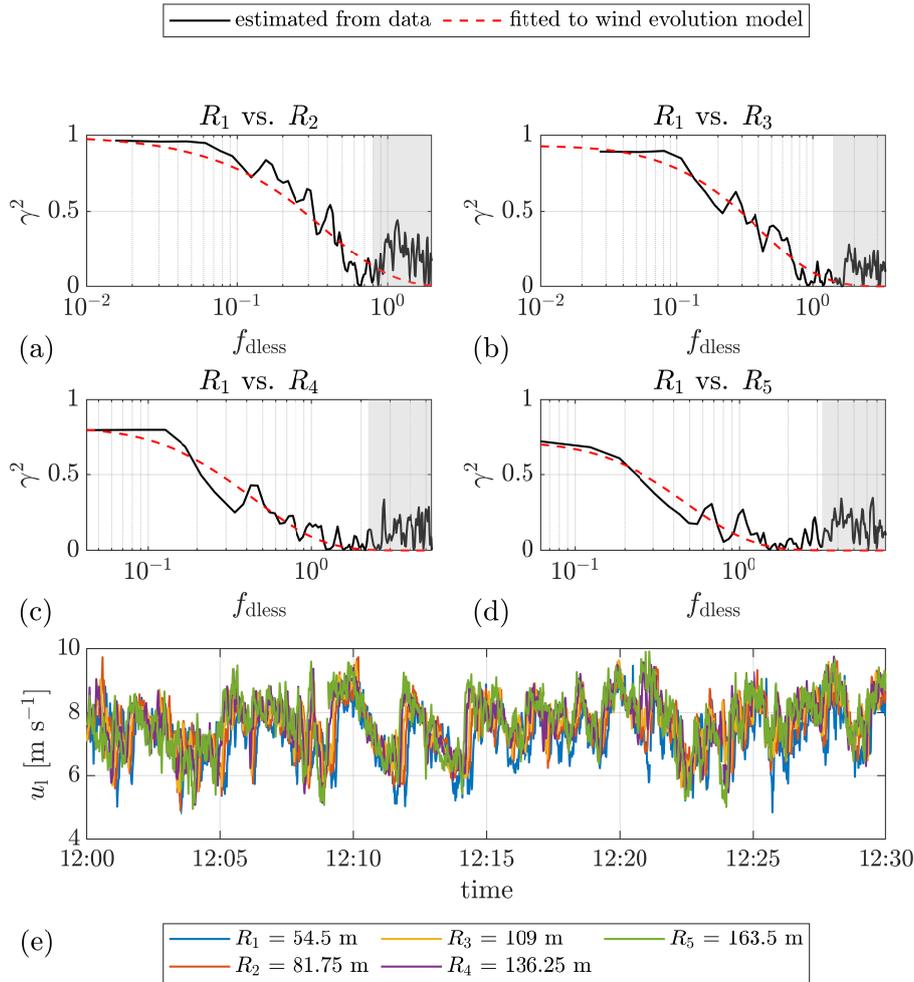


Figure 5. (a) – (d) Example plots of the estimated coherence between the lidar wind speeds measured at different range gates and the corresponding fitted curves. The separations between the corresponding range gates are 27.25 m, 54.5 m, 81.75 m, and 109 m, respectively. The shaded areas indicate the data filtered by the cut-off frequency 0.2 Hz. (e) Time series of the lidar wind speed. The mean lidar wind speed $U_1 = 7.3 \text{ m s}^{-1}$. U_1 ranges from 7.3 m s^{-1} to 7.7 m s^{-1} and the lidar measured turbulence intensity $I_{T,1}$ ranges from 0.10 to 0.12, for different range gates. Date: 07 Dec. 2013. Data source: LidarComplex.

675 ~~the~~ both distributions, two special post-processings are executed correspondingly. Firstly, because the lidar sampling rate of LidarComplex is approximately ~~as three times as three times~~ that of ParkCast, an artificial data set is made for LidarComplex by averaging every three data points of the original lidar data to simulate ~~the lidar data as if it were measured with a similar~~ ~~sampling rate as measurement at a sampling rate similar to~~ that of ParkCast, so that the distributions of ~~the~~ both measurements can be compared. The fitted probability density function (PDF) of the wind evolution model parameters determined with this data set are plotted as yellow dashed lines in Fig. 6 (a) and (b). The comparison between the fitted PDF of the original

data and that of the data with reduced sampling rate indicates that the lidar sampling rate only very slightly affects the wind evolution model parameters, or, perhaps more accurately, the estimated coherence. Hence, the different sampling rates ~~of the~~
680 ~~both measurements~~ do not account for the differences between the ~~both~~ cases observed in Fig. 6. Secondly, because of the limited measurement range of LidarComplex, the maximum separation between two range gates ~~only reaches~~ reaches only 109 m, while that of ParkCast reaches more than 700 m. To make them comparable, Fig. 6 (c) and (d) ~~merely show~~ show only the wind evolution parameters calculated from the coherence with separation ~~smaller than~~ below 120 m of ParkCast.

Apart from that, the measurements were carried out in different environments (onshore and offshore), at different times of the
685 year (which ~~would impact~~ impacts atmospheric stability), and have different wind speed and turbulence intensity distributions (see Fig. A1). Despite these differences, the ~~value~~ distributions of the wind evolution model parameters do have some common characteristics. First of all, the value ranges of ~~the~~ both wind evolution model parameters for ~~the~~ both measurements are similar; ~~;~~ a ranges mostly from 0 to 6 and b from 0 to 0.5. Values out of these ranges are less likely to happen, according to the measurements. Second, the values of a and b are found ~~out~~ to follow an inverse Gaussian distribution and a Gamma
690 distribution, respectively. These two PDFs are determined by fitting the histograms to all the PDFs supported by ~~MATLAB and~~ the MATLAB Statistics and Machine Learning Toolbox and searching for the one with the maximum likelihood. This is done using a tool called fitmethis⁵.

The corresponding fitted parameters of the PDFs (orange curves) are displayed in Table 3. It is interesting to observe that the peak of the probability density is located around ~~$a = 1.8$~~ $a = 1.8$ for the onshore ~~case~~ LidarComplex, while around ~~$a =$~~
695 ~~0.8~~ $a = 0.8$ for the offshore ~~case~~ ParkCast. Moreover, the medians of a are ~~approx. 2.0 and 1.5~~ approximately 2.0 and 1.5 for LidarComplex and ParkCast, respectively. The mean (see μ in Table 3) and median of a as well as its value of the peak location of the PDF of LidarComplex are all higher than that of ParkCast. This indicates that the coherence under similar separation generally decays faster in an onshore location than an offshore location. In terms of b, most of the values ~~is near 0~~ are near 0, and values higher than ~~0.1~~ 0.1 are not often observed. Therefore, the ~~y-axes~~ y-axes in Fig. 6 (b) and (d) are ~~shown in logarithmic~~
700 ~~manner~~ plotted logarithmically to make the ~~part of higher value of~~ higher-value part of b visible. However, ~~there is~~ b shows no significant difference ~~of b between both~~ between the two cases observed in the figure.

It is not yet possible to explain the physical relationship between the wind evolution model parameters and the above-mentioned PDFs and the physical meaning of the corresponding PDF parameters. To verify whether the ~~above-discussed~~
~~phenomena commonly exist~~ above-discussed phenomena commonly occur in wind evolution, further research involving more
705 different measurement campaigns is necessary. At this point, a hypothesis is made that the values of a and b might follow an inverse Gaussian distribution and a Gamma distribution, respectively. The corresponding PDF parameters ~~probably might~~ depend on the terrain types, on the one hand. It is not clear if the roughness length would be ~~the a~~ a suitable parameter to quantify the influence of the terrain type on the value distribution of wind evolution model parameters. To figure out a concrete relationship between the PDF parameters and the terrain types, again, it is necessary to involve more measured data gathered
710 from different terrain types. On the other hand, unfortunately, it is not yet possible to estimate to what extent the atmospheric

⁵Francisco de Castro (2020). fitmethis (<https://www.mathworks.com/matlabcentral/fileexchange/40167-fitmethis>). MATLAB Central File Exchange. Retrieved Jan 13, 2020.

stability would affect the distribution of the wind evolution model parameters because there was no sonic data available for ParkCast to ~~do the relevant research~~ inform the associated investigation until this work was finished.

Table 3. Parameters of the fitted probability density functions.

wind evolution model parameters	PDF	LidarComplex	ParkCast
a	inverse Gaussian distribution	$\mu = 2.07$	$\mu = 1.86$
	$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]$	$\lambda = 17.23$	$\lambda = 2.38$
b	Gamma distribution	$k = 0.42$	$k = 0.24$
	$f(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	$\theta = 0.18$	$\theta = 0.16$

Note: the notations μ, λ, k, θ are independent from the other notations in the table.

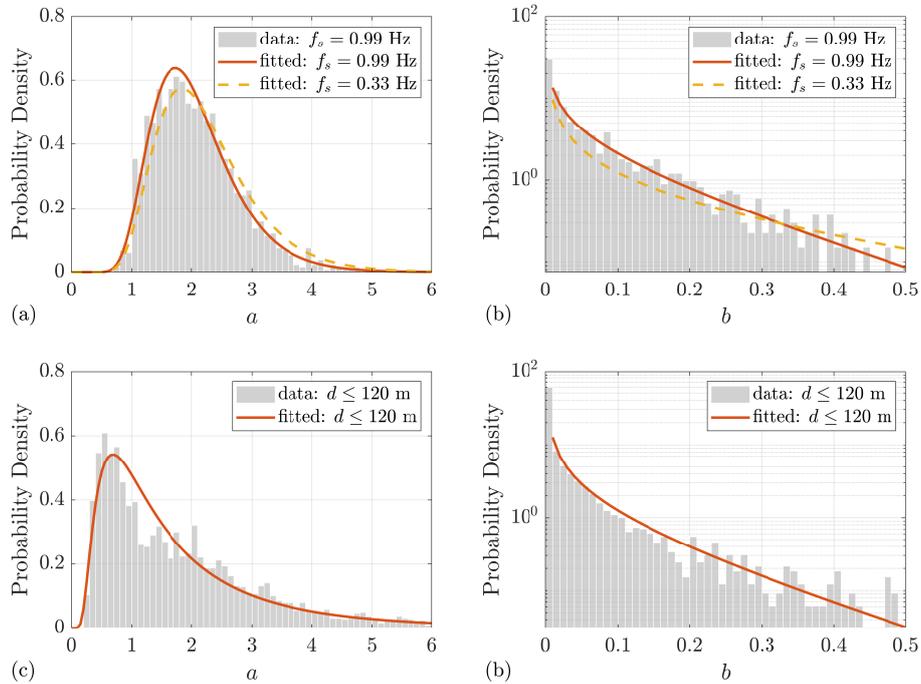


Figure 6. Distribution of wind evolution model parameters. (a) and (b): LidarComplex. (c) and (d): ParkCast. The curves show the corresponding fitted probability density function.

4.2 Dependency-Dependence of the Wind Evolution Model Parameters on Measuring-Separationmeasurement separation

715 Figure 7 shows the fitted curves of the estimated coherence of all pairings of the above-mentioned case-study-example-of
LidarComplex. The same color indicates the pairings have a common LidarComplex case-study example. Each color indicates
a particular range gate, while the same marker indicates the pairings have the same measuring separation. From the figure
it can be observed each marker indicates a particular measurement separation. The figure shows a very clear dependency
dependence of the fitted curve form on the measuring separation—The measurement separation — the curves with the same
720 marker are overlapped despite overlap despite having different range gates. This confirms that the coherence only depends
on the relative position but has nothing to do with the absolute position separation of the measurement points but not on their
positions, even though the measurement points are located in the induction zone of a wind turbine, which is defined as the range
of 2.5 rotor diameters in the wind turbine's induction zone (defined as within 2.5 rotor diameters on the inflow side of the wind
turbine). Since the curve offset is only-related-related only to the offset parameter b , obviously, b must strongly depends-on-the
725 measuring depend on the measurement separation. In addition, that all the fitted curves of the coherence are grouped together
proves-suggests it is reasonable to model the wind evolution based on the dimensionless frequency. Similar conclusions can be
drawn from the example plot of ParkCast (see Fig. A5), which proves that these conclusions are not accidental.

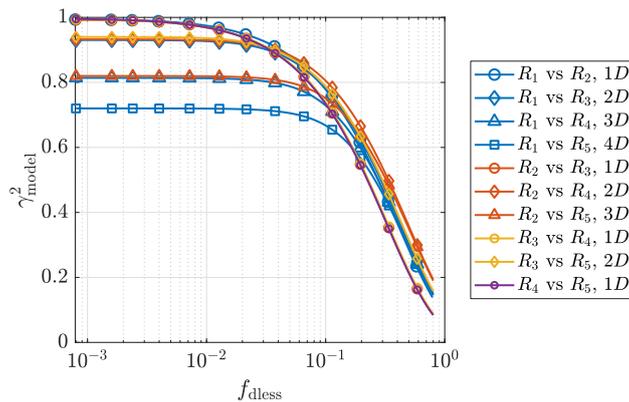


Figure 7. Fitted curves of the estimated coherence between the lidar wind speeds measured at different range gates. The range gate-gates
 R_1 to R_5 locate are located at 54.5 m, 81.75 m, 109 m, 136.25 m, and 163.5 m, respectively. $1D = 27.25$ m. The mean lidar wind speed
 $U_T = 7.3 \text{ m s}^{-1}$ U_1 ranges from 7.3 m s^{-1} to 7.7 m s^{-1} and the lidar measured turbulence intensity $I_{T,1}$ ranges from 0.10 to 0.12, for different
range gates. Date and time: 12-07 Dec. 2013, 12:00-12:30. Data source: LidarComplex.

To further study the dependency-dependence of the wind evolution model parameters on the measuring-measurement sep-
aration, the box plots of the wind evolution parameters, grouped by the measuring-separationsmeasurement separations, are
730 given in Fig. 8. Although the ranges of the measuring separation of the both measurements measurement separation from the
two measurement campaigns are very different, similar trends can still be observed from the box plots still show similar trends.
The decay parameter a shows a decreasing trend with increasing measuring-measurement separation. This decreasing trend of

a gradually stops ~~when the separation approaches 300 m~~ at a separation of about 300 m, as observed in Fig. 8 (c). The offset parameter b shows an increasing trend with ~~increasing measuring~~ separation. An ~~increasing increase in~~ b implies a ~~decrease of~~ the decreased offset of the coherence curve. This is consistent with the phenomena observed from Fig. 7 and Fig. A5.

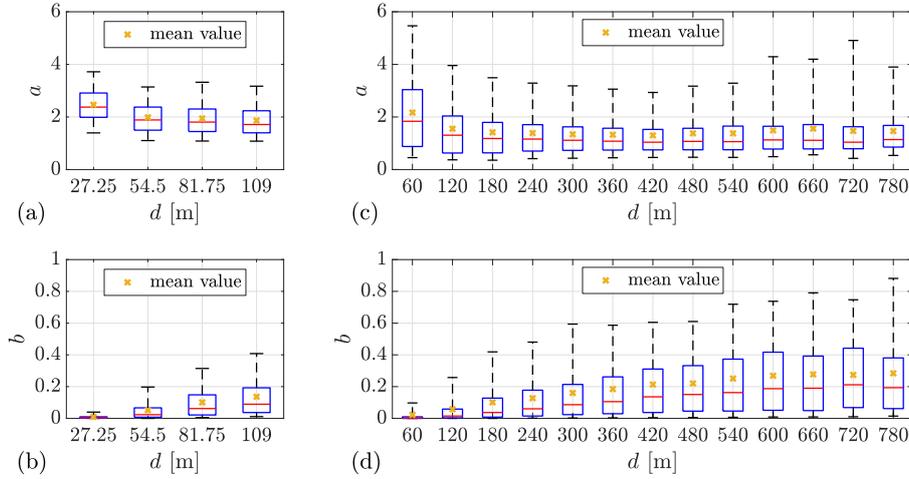


Figure 8. Box plots of the wind evolution model parameters grouped by the ~~measuring-measurement~~ separations d . (a) and (b): LidarComplex. (c) and (d): ParkCast. The bottom and top of the ~~boxes-boxes~~ indicate the first (25th percentile) and the ~~third~~ (75th percentile) quartilequartiles. The lower and upper whiskers show 5th (bottom) and 95th (top) percentiles. The red line in the middle indicates the median value. Minimum sample size is 50.

The decay of ~~the~~ coherence is supposed to result from the evolution of ~~the turbulence eddies with respect to the turbulence eddies depending on~~ travel time. The ~~dependency-dependence~~ of the decay parameter a on the ~~measuring-measurement~~ separation, or rather the travel distances, actually reveals the ~~dependency-dependence~~ of a on the travel time. Figure 9 shows the correlation between a and the travel time approximated by $\Delta t_{\max\text{corr}} - \Delta t_M$ of ParkCast. The fitted curve represents a negative correlation trend between them. This implies that the decay rate of the coherence decreases with ~~an~~ increasing travel time. The nonlinear least-squares fitting is done ~~through nonlinear least-squares fitting with using the~~ Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963; Moré, 1978).

5 Parameterization Model

This section first presents the training procedure of ~~the GPR model~~ GPR models with application of the ARD-SE kernel to select the suitable predictors in Sect. 5.1. ~~Followed~~ Following that is a discussion ~~about the predictor selection of the selected predictors~~ in Sect. 5.2, and an evaluation of the model performance ~~of the GPR models for the prediction of the wind evolution model parameters~~ in Sect. 5.3.

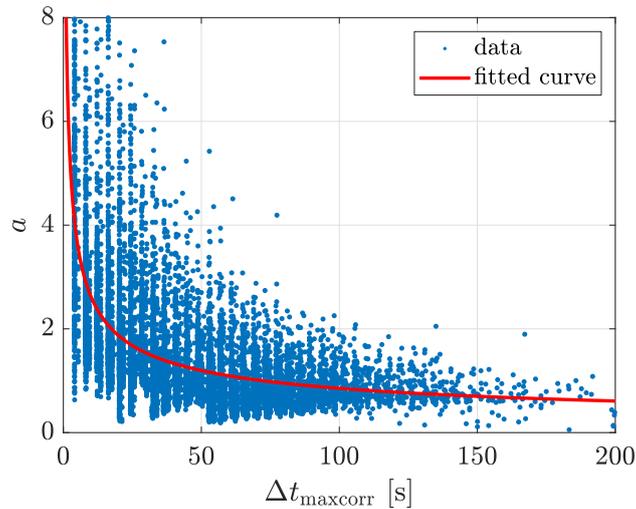


Figure 9. Correlation between the decay parameter a and the travel time approximated by $\Delta t_{\max\text{corr}}$. The equation of the fitted curve: $a = 8.06 \cdot \Delta t_{\max\text{corr}}^{-0.49}$. Data source: ParkCast.

5.1 Model Training

The model training is a two-step process. In the first step, all the potential predictors are included in a preliminary model training to determine the characteristic length scale σ_m for each predictor (see Eq. 27). As explained in Sect. ??, a relatively large $\log(\sigma_m)$ indicates that the corresponding predictor is less relevant to the model, and thus is not a suitable or necessary predictor for the model, and vice-versa. Figure 10 illustrates a comparison among the $\log(\sigma_m)$. In the second step, the predictors are selected according to different preset limits of the $\log(\sigma_m)$ considering different cases of application or data availability, e.g. if only lidar data is available or if sonic data is also available. The models are trained again using the corresponding selected predictors with application of a 5-fold cross-validation. The corresponding RMSE (see Eq. 29) and R^2 (see Eq. 30) are calculated and used as criteria for the evaluation and comparison of the model performance.

Comparison of the relative importance of the predictors. (a) ParkCast, lidar data. (b) LidarComplex, lidar data. (c) LidarComplex, sonic data.

The initial settings of the initial settings for GPR model training is are listed in Table 4. The setting of "exact GPR" means 'exact GPR' setting means that a standard GPR is applied in the fitting and prediction process. In contrast, the otherwise GPR can be approximated using different methods to reduce the computational computation time for large amount amounts of training data. The initial values of σ_n , σ_f , and σ_m listed in the table are just used to initiate the training. The final values are determined by process, and their final values will be estimated from the training data. The standardization is done by the GPR algorithm. The training data is standardized by centering and scaling the data of each predictor by its mean and standard deviation, respectively, which gives the standard scores (also called z-scores) z-scores (Kreyszig, 1979; Mendenhall and Sincich, 2007) of the predictor data. When the standardization is activated, the training is done using the standardized predictor data.

Table 4. Initial settings of GPR model training.

<u>Hyperparameter</u>	<u>Setting</u>
Basis function	constant
Kernel function	ARD-SE
Fitting method	exact GPR
Prediction method	exact GPR
Initial value of σ_n	standard deviation of observed responses <u>target values</u>
Initial value of σ_f	standard deviation of observed responses <u>target values</u>
Initial value of σ_m	10
Standardization	true

5.2 Discussion of Predictor Selection

770 In the predictor selection, two different situations of data availability are considered: only using variables calculated with lidar data as predictors (in both of LidarComplex and ParkCast available) and only using variables calculated with sonic data (only in LidarComplex available). The main focus of this study is to evaluate the possibility to predict the wind evolution only using lidar data. Including the situation of only using sonic data aims to provide a comparison. Training the model is a two-step process. In the first step, all the potential predictors are included in a preliminary training to determine the characteristic length scale σ_m for each predictor (see Eq. 27). Figure 10 illustrates a comparison among the $\log(\sigma_m^{-2})$ of all potential predictors. As explained in Sect. 2.6, the larger $\log(\sigma_m^{-2})$ is, the more important and useful the corresponding predictor is for a GPR model, and thus this predictor should be selected. In the second step, new GPR models are trained only with the selected predictors, applying a 5-fold cross-validation to evaluate the model performance, using RMSE (see Eq. 29) and R^2 (see Eq. 30) as criteria.

780 Table 5 shows the different predictor combinations and the corresponding limit values of the $\log(\sigma_m)$ for the both wind evolution model parameters and for the both measurements. The corresponding Table 5 displays the predictors selected according to different lower limits of $\log(\sigma_m^{-2})$ under different measurement campaigns (LidarComplex or ParkCast), different data availability (whether sonic data is available), and different targets (a or b). R^2 and the RMSE of the 5-fold cross-validation are also listed. The bold text indicates the recommended predictor combinations for each situation. The recommendation principle is achieving as high model performance as possible with as few predictors as possible as well.

785 The overall performance of the GPR model is satisfactory — in all situations, the R^2 of the recommended cases are over 0.65 and for the best case can even reach 0.8 (see case 14). These results are much better than that of the preliminary study (Chen, 2019), especially the prediction accuracy of the offset parameter b has been significantly improved. This is attributed to the application of the ARD-SE kernel in this study, whereas in the preliminary study, kernel functions with a common

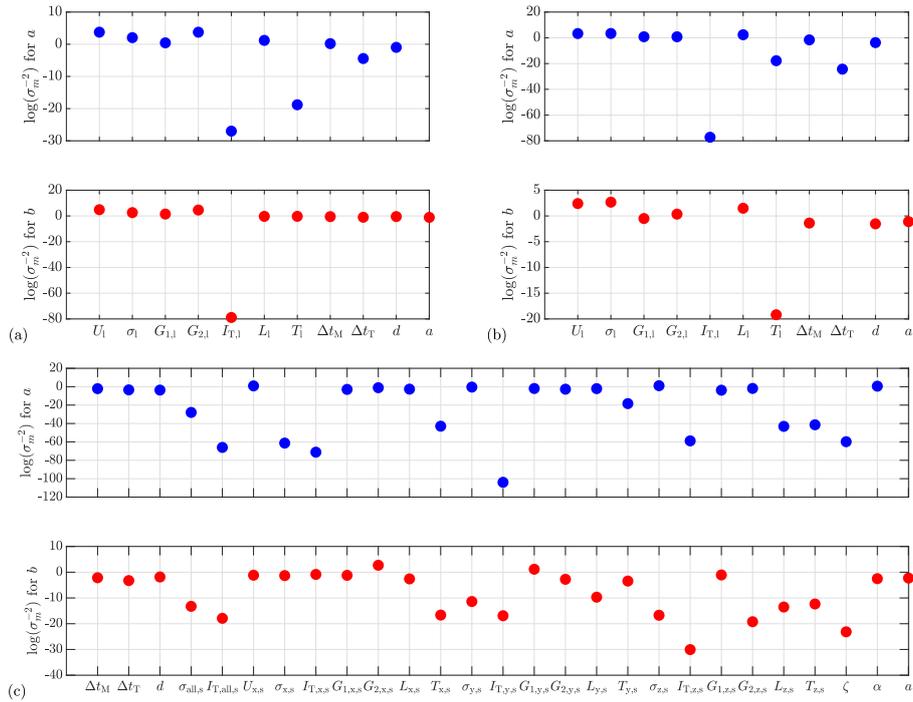


Figure 10. Comparison of the relative importance of predictors. (a) ParkCast, lidar data. (b) LidarComplex, lidar data. (c) LidarComplex, sonic data. $\log(\sigma_m^{-2}) = -\infty$ is not displayed.

length scale for predictors were tried. Moreover, for the cases of using lidar data, the same predictors are selected for the both measurements despite the very different measurement conditions, which proves the corresponding predictor combinations are not accidental. Therefore, it is believed that the ARD-SE kernel is able to select reasonable predictors for the GPR model. As shown in general, the more relevant predictors are involved in the model, the more accurate predictions the model can make. However, using more predictors entails a larger training data set and thus a longer model training time. On the other hand, it might also reduce the applicability of the model because predictions can only be made when all predictors are consistently available and reliable. The trade-off between these factors must be considered in predictor selection, and it is aimed to achieve relatively high model performance with as few predictors as possible. The bold text in Table 5, the predictors with $\log(\sigma_m)$ lower than -1.0 indicates the recommended predictor combinations for each situation based on these considerations. The predictors with $\log(\sigma_m^{-2}) > -2$ are generally essential for the model. The limit of $\log(\sigma_m)$ can be set to 0 if the

Let us take the situation of using lidar data from LidarComplex to predict a as an example to explain the process of predictor selection (see Fig. 10 (a) top and the first block in Table 5). Firstly, since $\log(\sigma_m^{-2})$ of $I_{T,l}$ and T_l are much smaller than the others, it is not necessary to consider these two predictors, and the lower limit of $\log(\sigma_m^{-2})$ can be initially set to -4 (see Table 5: Case 1). Then, try to increase the lower limit of $\log(\sigma_m^{-2})$ step by step, e.g. first to -2 (see Table 5: Case 2) and then to 0 (see Table 5: Case 3), to further reduce the number of predictors needs to be reduced while the model performance should be

Table 5. Results Summary of the predictors selected according to different lower limits of $\log(\sigma_m^{-2})$ under different measurement campaigns, different data availability, and different targets. R^2 and RMSE are obtained from a 5-fold cross-validation of the model **cross-validation** trained with the respective combination of predictors. The bold text indicates the recommended predictor combinations.

Measurement	Target	Case	$\log(\sigma_m) - \log(\sigma_m^{-2})$	Predictors
LidarComplex lidar data	a	1	$\langle -1.0 \rangle -4$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, G_{1,j}, G_{2,j}, L_1, T_1, \Delta t_{\text{Taylor}}, \Delta t_M, d$
		2	$\langle 0 \rangle -2$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, T_1, \Delta t_{\text{Taylor}}, G_{1,j}, G_{2,j}, L_1, \Delta t_M$
		3	$\langle -1.0 \rangle > 0$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, G_{1,j}, G_{2,j}, L_1, T_1, \Delta t_{\text{Taylor}}, d, \alpha$
		4	$\langle 0.1 \rangle$	$U_1, \sigma_1, L_1, \Delta t_M, d$
		5	\sim	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, T_1, G_{1,j}, G_{2,j}, L_1, \Delta t_T, d$
		5-6	$\langle -0.8 \rangle$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, G_{1,j}, G_{2,j}, L_1, \Delta t_M, d, \alpha$
LidarComplex lidar data	b	6	$\langle -1.0 \rangle -2$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, G_{1,j}, G_{2,j}, L_1, \Delta t_{\text{Taylor}}, \Delta t_M, d$
		7*8	≥ -1	$U_1, \sigma_1, G_{1,j}, G_{2,j}, L_1$
		9	-	$U_1, \sigma_1, L_1, \Delta t_{\text{Taylor}}, \Delta t_M, d$
		8-10	$\langle -1.0 \rangle$	$U_1, \sigma_1, G_{1,j}, G_{2,j}, L_1, \Delta t_T, d$
		9-11	$\langle -1.0 \rangle$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, T_1, G_{1,j}, G_{2,j}, L_1, \Delta t_M, d, \alpha$
		10-12	$\langle -1.0 \rangle$	$U_1, \sigma_1, \tilde{\mu}_{3,1}, \tilde{\mu}_{4,1}, T_1, G_{1,j}, G_{2,j}, L_1, \Delta t_M, d, \alpha$
LidarComplex sonic data	a	11	$\langle -2.0 \rangle -4$	$U_{x,s}, \sigma_{x,s}, \tilde{\mu}_{3,x,s}, \tilde{\mu}_{4,x,s}, G_{1,x,s}, G_{2,x,s}, L_{x,s}, \Delta t_{\text{Taylor}}, \sigma_{y,s}, \tilde{\mu}_{3,y,s}, G_{1,y,s}, G_{2,y,s}, L_{y,s}, \sigma_{z,s}, G_{1,z,s}, G_{2,z,s}, \Delta t_M, d, \alpha$
		12-14	$\langle -0.5 \rangle -2$	$U_{x,s}, \sigma_{x,s}, \tilde{\mu}_{3,x,s}, \tilde{\mu}_{4,x,s}, G_{2,x,s}, \sigma_{y,s}, G_{1,y,s}, \sigma_{z,s}, G_{2,z,s}, \alpha$
		13-15	$\langle 0 \rangle > 0$	$U_{x,s}, \sigma_{x,s}, \sigma_{z,s}, \alpha$
LidarComplex sonic data	b	14	$\langle -1.0 \rangle > -3$	$\tilde{\mu}_{3,x,s}, \tilde{\mu}_{4,x,s}, U_{x,s}, I_{T,x,s}, G_{1,x,s}, G_{2,x,s}, L_{x,s}, \Delta t_{\text{Taylor}}, \sigma_{y,s}, I_{T,z,s}, \tilde{\mu}_{4,z,s}, L_{z,s}, G_{1,y,s}, G_{1,z,s}, \Delta t_M, d, \alpha, a$
		15-17	$\langle -0.5 \rangle -2$	$\tilde{\mu}_{3,x,s}, L_{x,s}, \alpha, U_{x,s}, I_{T,x,s}, G_{1,x,s}, G_{2,x,s}, G_{1,y,s}, G_{1,z,s}, d$
		18	≥ -1	$I_{T,x,s}, G_{2,x,s}, G_{1,y,s}$
ParkCast lidar data	a	19	≥ -1	$U_1, \sigma_1, G_{1,j}, G_{2,j}, L_1, \Delta t_M, d$
		20	> 0	$U_1, \sigma_1, G_{1,j}, G_{2,j}, L_1, \Delta t_M$
ParkCast lidar data	b	21	≥ -1	$U_1, \sigma_1, G_{1,j}, G_{2,j}, L_1, \Delta t_M, d$
		22	> 0	$U_1, \sigma_1, G_{1,j}, G_{2,j}$

Notes: Cases 4–6 and Cases 9–12 are selected for comparison with Case 1 and Case 7 for different purposes, respectively. Case 4 and Case 9: to examine the effect of introducing G_1 and G_2 as predictors. Case 6 and Case 11: to examine the effect of having α available. Case 12: to examine the effect of introducing a as a predictor for b .

kept acceptable. The resulting models are evaluated to determine whether it is appropriate to remove these predictors. For example, the comparison between Case 1 and Case 2 shows that removing d almost does not affect the model performance in this situation, with R^2 decreasing only slightly from 0.70 to 0.69. However, further abandonment of Δt_M significantly reduces

the prediction accuracy, reducing R^2 more substantially from 0.69 in Case 2, to 0.59 in Case 3. Therefore, it is no longer proper to remove further predictors, and the predictor combination in Case 2 is recommended.

The following points about predictor selection are worth discussing:-

810 5.2 Discussion of Selected Predictors

Feature selection is not only a tool to select suitable predictors for a machine learning model, but also could shed some light on intrinsic relationships among data. Here are some discussions about the selected predictors, to provide some insights into possible correlations between wind evolution and these predictors.

Selection between two related variables. In the preliminary training, two pairs of related variables are intentionally involved at the same time: standard deviation σ and turbulence intensity I_T , integral time scale T and integral length scale L . Only 815 It is only necessary to select one of the two related variables needs to be selected as predictor if necessary (if determined to be relevant), because they can be converted to each other into each other (given U). In terms of σ and I_T , it is surprising to notice that the GPR model shows models show a preference for σ rather than I_T , although I_T is more commonly used in data analysis and simulation in wind energy. The only exception is the situation of using sonic data from LidarComplex to predict 820 b (Cases 16–18). It is possible that GPR generally tends to select fundamental variables (directly calculated from measured data) instead of derived variables (calculated from other variables). However, the situation becomes complicated in terms of selection becomes complicated for T and L . The GPR model shows a slight preference for T in most of the cases of using lidar data except in the case 6–8 where In some situations, L is clearly more preferred. Considering that, e.g. $\log(\sigma_m^{-2})$ of L is actually approximated by the multiplication obviously higher than $\log(\sigma_m^{-2})$ of T and the mean wind speed (see Eq. (18)) 825 and the relationship between σ and I_T is also similar, it is possible that the GPR model generally tends to select the variable calculated from the data but not from the other variables in Fig. 10 (a) top and (b). In the other situations, $\log(\sigma_m^{-2})$ of L and $\log(\sigma_m^{-2})$ of T show similar values. For consistency, we decided to select L for all cases whenever L is determined to be relevant.

Introducing higher-order wind statistics as predictors. In literature, the higher-order statistics skewness and kurtosis of the 830 wind speed are not considered in the study of the wind evolution. So far, skewness G_1 and kurtosis G_2 of wind speed have not been considered in wind evolution research. However, it is interesting to observe that they worth noting that both are selected as predictors for all the cases of using lidar data besides the mean and the standard deviation. The comparison between the case 6 and 7 proves that introducing skewness and kurtosis in all cases except Case 15, despite different measurement sites and devices. Case 4 and Case 9 are aimed at examining the effects of G_1 and G_2 on the prediction of a and b , respectively, with G_1 835 and G_2 removed in comparison to Cases 1 and 7. Case 4 and Case 9 show much worse prediction accuracy, with $R^2 = 0.53$ in Case 4 compared to $R^2 = 0.70$ in Case 1 and $R^2 = 0.46$ in Case 9 compared to $R^2 = 0.70$ in Case 7. This comparison confirms that G_1 and G_2 are essential for predicting wind evolution when using lidar data, and introducing G_1 and G_2 as predictors can indeed improve the prediction accuracy of the wind evolution. This also implies that the skewness and the kurtosis probably significantly improve the models, despite uncertainties contained in their estimated values from measured data (see Sect. 2.5). 840 This implies that G_1 and G_2 might contain additional information which can describe the state of turbulence more concretely.

For example, given the same could distinguish different states of turbulence given a particular mean wind speed and standard deviation, a positive or a negative skew could indicate two different states of the turbulence, which could further be associated with other conditions: turbulence intensity, and this ‘different state’ might be relevant to wind evolution.

845 **Different approximations of travel time.** Δt_M and Δt_T are two different approximations of travel time. Although Δt_M is expected to be more predictive than Δt_T , Δt_T is still involved in the model training because, in application, it is easier to calculate Δt_T than Δt_M . Cases 5 and 10 are selected to compare with Cases 1 and 7, respectively, to examine the different effects of Δt_T and Δt_M on the GPR models. The respective values of R^2 show that replacing Δt_M with Δt_T only slightly decreases the prediction accuracy. Therefore, for a simpler calculation of travel time, Δt_T can be used as a predictor instead.

850 **Effect of misalignment angle.** As discussed in Sect. 2.5, misalignment angle α is supposed to be an important predictor for the prediction of the horizontal coherence. In Cases 13–15, where sonic data from LidarComplex is used to predict a , α shows a high relevance with $\log(\sigma_m^{-2}) > 0$. However, for the prediction of b using sonic data (Cases 16–18), removing α from predictors does not influence the prediction accuracy much, especially when comparing Case 16 and Case 17, with $R^2 = 0.80$ and $R^2 = 0.78$, respectively. These results indicate that α is essential for the prediction of a but not relevant for predicting b .

855 In addition, α is introduced in the prediction using lidar data (Case 6 and Case 11) as well to examine its effect, although α is actually not available when only using a lidar in staring mode. As mentioned in Sect. 3.1, α is approximated by the deviation between the yaw position of the turbine and mean wind direction taken on the meteorological mast. Cases 6 and 11 both show better prediction accuracy than Cases 1 and 7, with $R^2 = 0.76$ and $R^2 = 0.81$, respectively, despite the uncertainties in the approximation of α . This means that if α were available, the prediction accuracy of the models trained with lidar data could be further improved. As mentioned earlier, α could be made available e.g. by deploying a multi-beam lidar.

860 **Introducing one of the targets as a predictor for the other.** ~~Since the two wind evolution parameters~~ According to the wind evolution model (Eq. (8)), a and b jointly determine the shape ~~of the wind evolution model, there is~~ and the position of the modeled coherence, and thus they have a certain correlation ~~between them with each other~~. Introducing one of them as a predictor for the other may improve its prediction accuracy. ~~The comparison of case 9 and 10~~ Case 12, with $R^2 = 0.74$, compared to $R^2 = 0.70$ in Case 7, confirms that introducing a as a predictor for b can ~~indeed increase the R^2 of the~~ help with ~~the~~ prediction of b . This means it could be a good idea to predict the wind evolution model parameters successively rather than ~~separately in parallel~~. This concept is not yet fully studied in this work, and thus ~~case 9 is not considered as~~ Case 12 is not presented as a recommendation. To prove its applicability, it is necessary to investigate which wind evolution model parameter should be first predicted, and how the prediction uncertainty ~~of the first one would be propagated in the first parameter would propagate~~ to the second one.

870 **Prediction using sonic data.** ~~As mentioned above, the research on the cases of~~ Additional research on using sonic data as predictors ~~is involved to provide a comparison. Figure 11 illustrates a comparison between~~ aims to provide some insights into whether it is worth involving sonic data in wind evolution prediction when available. When comparing the model performance of ~~the recommended cases of using lidar data and sonic data~~ using lidar data and sonic data from LidarComplex, Case 13—the best case of using sonic data to predict a —shows a higher prediction accuracy ($R^2 = 0.83$) than Case 6—the best case of using lidar data given α available ($R^2 = 0.76$). However, Case 13 needs many more predictors than Case 6, whereas Case 14 and

Case 15, with fewer predictors, do not show any advantage in prediction accuracy. For predicting b , Case 16—the best case of using sonic data ($R^2 = 0.80$), does not outperform Case 11, the best case of using lidar data given α available ($R^2 = 0.81$). It must be emphasized that the ultrasonic anemometer is installed on a ~~met mast located at meteorological mast located~~ 295 m away from the lidar. There must be a deviation between the sonic data and the ~~data of true values in~~ the wind field where the coherence is estimated. ~~Despite this, the best cases of~~, which reduces the prediction accuracy when using sonic data as predictors (case 11 and 14) show a higher prediction accuracy, R^2 of about 0.8, in comparison to that of. Figure 11 illustrates a comparison between the model performance of the recommended cases of using lidar data (case 6 and 9). This implies that the upper limit of the prediction accuracy will be higher if sonic data is used. Moreover, case 13 shows a slightly better result than that of case 6 and sonic data.

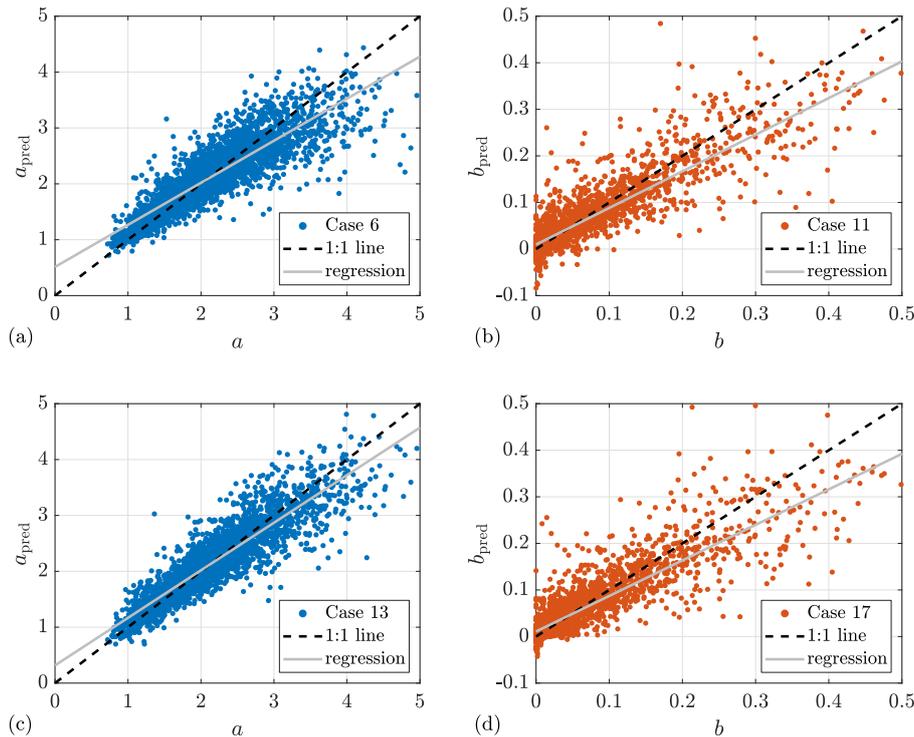


Figure 11. Comparison of prediction performance of models using lidar data and sonic data from LidarComplex. (a) and (b): lidar data. (c) and (d): sonic data. The subscript ‘pred’ indicates the predicted values.

885 Interestingly, Case 15 can achieve the same predictive accuracy as Case 1, with only three predictors: the mean and standard deviation of the longitudinal component mean wind speed $U_{x,s}$ and $\sigma_{x,s}$, and the standard deviation of the vertical wind component $\sigma_{z,s}$. Think of the physical consideration of Kristensen (1979). The Kristensen’s model assumes that the wind evolution is determined by two probabilities: one is the probability that the eddy decays within the travel time; the other is the probability that the eddy diffuses in, and the misalignment angle α . In fact, $\sigma_{z,s}$ is determined to be the most important

890 predictor by the transversal direction. The former could be associated with $U_{x,s}$ and $\sigma_{x,s}$, while the latter with $\sigma_{z,s}$. ARD-SE kernel, having the maximum value of $\log(\sigma_m^{-2})$. This might imply a possible correlation between wind evolution and vertical convection.

Influence of atmospheric stability. We initially intended to study the influence of atmospheric stability using a dimensionless height ζ as the stability parameter (see Sect. 2.5). However, very surprisingly, ζ is not selected as a relevant predictor in any cases, and $\log(\sigma_m^{-2})$ is quite small compared to the others (see Fig. 10 (c)). In the end, we found that the stability happens to be mostly neutral during the chosen measurement in LidarComplex. This could be an explanation for this predictor recombination: the reason for ζ not being selected as a predictor. Therefore, it is not possible to analyze the influence of atmospheric stability on wind evolution in this study.

Comparison of prediction performance of models using lidar data and sonic data of LidarComplex. The subscript "pred" indicates the predicted values.

5.3 Model Evaluation

As mentioned above, the shown in Table 5, R^2 of all recommended cases range from 0.67 to 0.83. These results are much better than that of the preliminary study (Chen, 2019); in particular, the prediction accuracy of the parameterization model using lidar data as predictors is satisfactory, showing an R^2 at least over 0.65. Section 5.2 already covers some contents of model evaluation, from the perspective of the prediction of offset parameter b has been significantly improved. This is mainly owing to the use of the wind evolution model parameters. In this section, model evaluation is further discussed from other perspectives. ARD-SE kernel, which can help to select predictors reasonably and give different weights to predictors according to their relevant importance for the prediction, whereas kernel functions with a common length scale for predictors were applied in the preliminary study.

910 Firstly, it is interesting to visualize the prediction errors of a and b are quantified with the respective RMSE between their predicted and observed values. But in fact, the shape and position of the predicted coherence determined by both parameters together is the final prediction goal. And the corresponding prediction errors will eventually appear as the deviation between the predicted curve and its estimated curve due to the prediction errors of a and b .

To intuitively display how the prediction errors affect the shape and the position of the predicted coherence in the frequency domain, combining the prediction of the both wind evolution model parameters. Figure Fig. 12 shows the predicted coherence and the corresponding 95% confidence interval by the GPR models using lidar data (case for the example case from LidarComplex. For the example prediction with lidar data in Fig. 12 (a), the prediction of a and b is made by the GPR models in Cases 6 and 10)11, respectively. And for the example prediction with sonic data in Fig. 12 (b), the prediction of a and sonic data (case b is made by the GPR models in Cases 13 and 14)for the case study example of LidarComplex17, respectively. The predicted coherence and the 95% confidence interval are reconstructed through by putting the predicted value values of a and b and their lower and upper bounds of the 95% confidence interval into the wind evolution model (Eq. (58)). From the figure, it can be observed that the prediction is very good for this example because the predicted coherence is almost overlapped with the one estimated from the measured data, and the 95% confidence interval is quite narrow.

925 To show the prediction errors in a more general sense, the RMSE interval is additionally indicated as shaded areas in Fig. 12. The lower and upper bounds of the RMSE interval are determined with

$$\gamma_{\text{model,lb}}^2(f_{\text{dless}}) = \exp \left[-\sqrt{(a_{\text{pred}} + \Delta a)^2 \cdot f_{\text{dless}}^2 + (b_{\text{pred}} + \Delta b)^2} \right] \quad (31)$$

and

$$\gamma_{\text{model,ub}}^2(f_{\text{dless}}) = \exp \left[-\sqrt{(a_{\text{pred}} - \Delta a)^2 \cdot f_{\text{dless}}^2 + (b_{\text{pred}} - \Delta b)^2} \right], \quad (32)$$

930 respectively, where a_{pred} and b_{pred} are the predicted values of a and b , and Δa and Δb are the respective RMSE. The narrow RMSE interval shows that the GPR models perform overall well in the prediction of wind evolution.

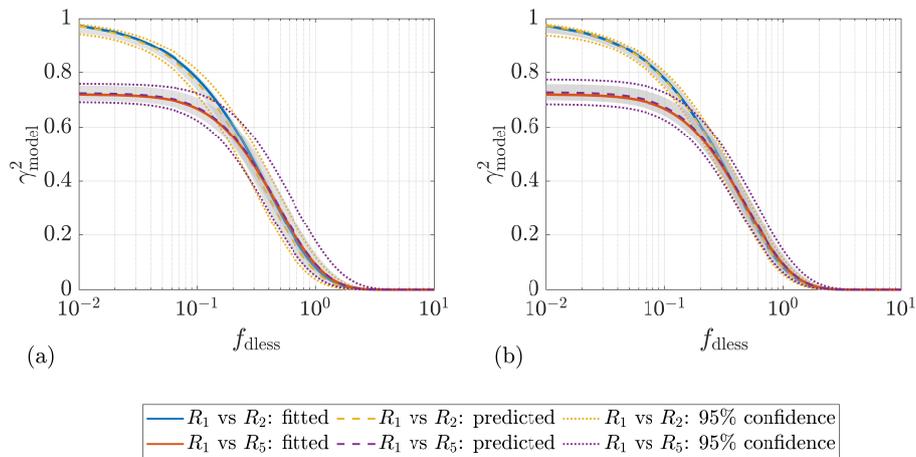


Figure 12. Example predicted coherence with 95% confidence interval for two different measuring-measurement separations of LidarComplex. (a) Prediction with lidar data: predicted value of a from case Case 6 and that of b from case 10. Case 11. (b) Prediction with sonic data: predicted value of a from case Case 13 and that of b from case 14. Case 17. The shaded areas indicate the RMSE interval. The input predictor data and the estimated coherence are from the case study example of LidarComplex: 12-07 Dec. 2013, 12:00-12:30. The mean lidar wind speed U_1 ranges from 7.3 m s^{-1} to 7.7 m s^{-1} and the lidar measured turbulence intensity $I_{T,1}$ ranges from 0.10 to 0.12, for different range gates.

935 Secondly Moreover, it is important to check if the prediction errors of the models are relevant to the values of the predictors. Taking the models trained with the lidar data of LidarComplex (case from LidarComplex (Case 6 and 10) as Case 11) as an example, Fig. ?? and ?? illustrate 13–16 show the box plots of the prediction errors, defined as the deviation between the predicted target value and the real one and the observed values of targets, with respect to the values of the predictors. The histograms of the predictor values are plotted below the box plots correspondingly. The x-axes x-axes of the box plots correspond the upper limit bound of the respective bin in the histograms. For example, in Fig. ??-13 (a), the first box labelled with "4" means it is plotted with the prediction errors of the samples attributed to the mean wind speed range of $3\text{--}4 \text{ m s}^{-1}$.

To avoid accidental conclusions, there is a minimum requirement of the sample size sample size requirement of 50 for the box plots. The

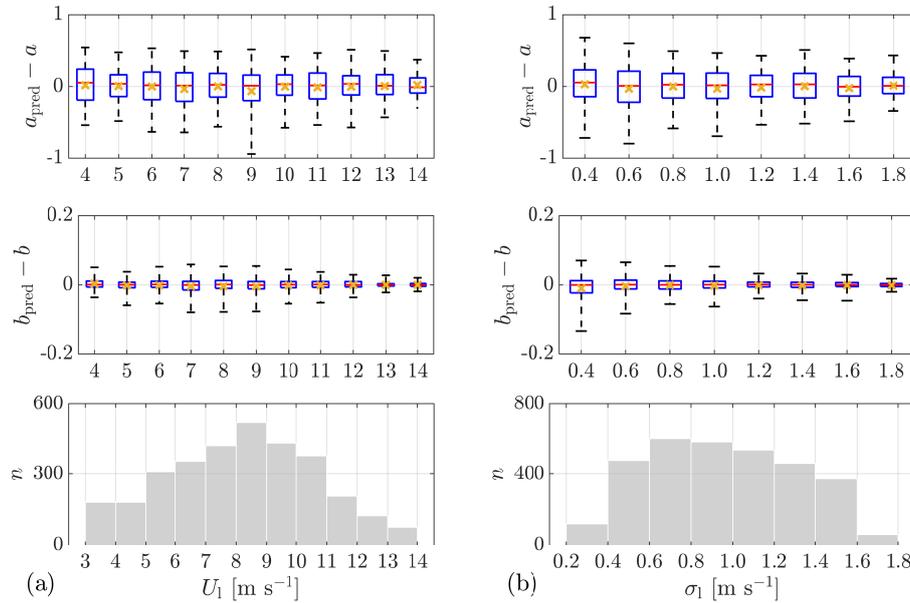


Figure 13. Prediction errors of a (Case 6) and b (Case 11) from LidarComplex with respect to the values of predictors. (a) Lidar measured mean wind speed U_1 . (b) Standard deviation of lidar measured wind speed σ_1 . n is sample size. The bottom and top of the boxes indicate the first and the third quartiles, i.e. 25th and 75th percentile, respectively. The lower and upper whiskers show 5th and 95th percentiles. The red line and the yellow cross in the middle indicate the median and mean value, respectively.

940 The box plots indicate data within the first and the third quartiles (i.e. 25th and 75th percentile) and represent the main part of the data, whereas whiskers show the tails of the distributions of the prediction error data indicating extreme values. In Fig. 13–16, it can be observed that the boxes of the prediction errors of a and b are all centered around 0 and no obvious relevance between the error and values of any of the predictors is indicated in the both figures. For some cases, the range of the whiskers is relatively large because of a small sample size, quite narrow and centered around 0, indicating small prediction errors for the majority of samples. That the boxes are centered around 0, as well as the median and mean values (indicated as red lines and yellow crosses, respectively), means that there is no systematic error with respect to predictor values. In the boxplots for the prediction errors of a , the ranges of boxes and whiskers do not show obvious relevance to predictor values except for small travel time and measurement separation. The large range of the box and whiskers of the first box in Fig. 15 (b) and that of the first box in Fig. 16 (a) implies that the prediction of a is likely more uncertain for small travel time and measurement separation

945 (both are related to some extent). The ranges of boxes and whiskers of the prediction errors of b show some relevance to the values of standard deviation, skewness, travel time, and measurement separation. In Fig. 13 (b), a clear trend can be observed, that the ranges of the boxes and whiskers decrease with the values of standard deviation, indicating that the prediction of b might be better for high turbulence. A similar trend can be observed in Fig. 14 (a), meaning that the prediction of b might be

950

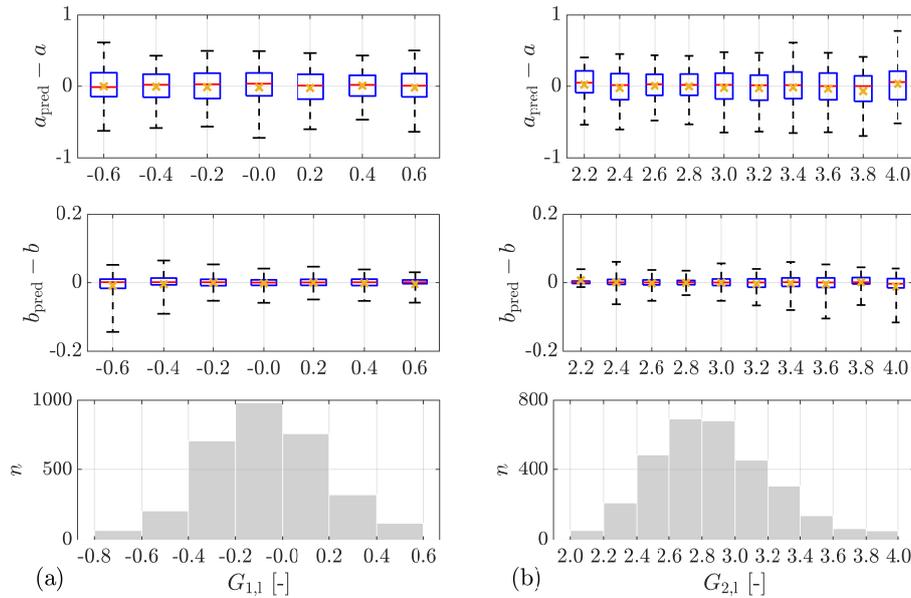


Figure 14. Prediction errors of a (Case 6) and b (Case 11) from LidarComplex with respect to the values of predictors. (a) Skewness of lidar measured wind speed $G_{1,1}$. (b) Kurtosis of lidar measured wind speed $G_{2,1}$. n is sample size. The bottom and top of the boxes indicate the first and the third quartile, i.e. 25th and 75th percentile, respectively. The lower and upper whiskers show 5th and 95th percentiles. The red line and the yellow cross in the middle indicate the median and mean value, respectively.

955 better under the circumstance of negative skewness (longer left tail) than that of positive skewness (longer right tail). In Fig. 15 (b) and Fig. 16 (a), the ranges of boxes and whiskers get larger with travel time and measurement separation, implying that the prediction errors of b increase with travel time and measurement separation.

Prediction error of a of case 6 with respect to the values of the predictors. n is sample size. The bottom and top of the boxes indicate the first and the third quartile, i.e. 25th and 75th percentile, respectively. The lower and upper whiskers show 5th and 95th percentiles. The red line and the yellow cross in the middle indicate the median and mean value, respectively.

960 Prediction error of b of case 10 with respect to the values of the predictors. n is sample size. The bottom and top of the boxes indicate the first and the third quartile, i.e. 25th and 75th percentile, respectively. The lower and upper whiskers show 5th and 95th percentiles. The red line and the yellow cross in the middle indicate the median and mean value, respectively.

As a result, it is proven that the Gaussian process regression is capable of achieving an accurate parameterization model for the wind evolution. However, it is worth emphasizing that the performance of any regression model is only possible to be can be only as good as the quality of the training data. No matter what kind choice of regression model cannot eliminate the noisy term in the can eliminate noise from the training data. And the noisier the training data is, the more uncertainties the prediction of the regression model will contain. A good data source is always very essential for training a good regression model.

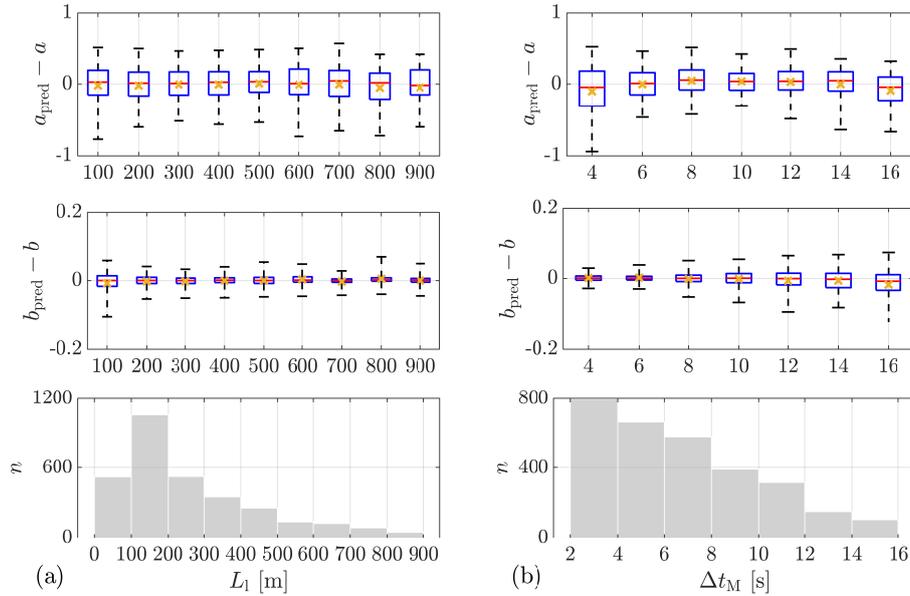


Figure 15. Prediction errors of a (Case 6) and b (Case 11) from LidarComplex with respect to the values of predictors. (a) Integral length scale of lidar measured wind speed L_1 . (b) Time lag determined by the peak of maximum cross-correlation Δt_M . n is sample size. The bottom and top of the boxes indicate the first and the third quartiles, i.e. 25th and 75th percentile, respectively. The lower and upper whiskers show 5th and 95th percentiles. The red line and the yellow cross in the middle indicate the median and mean value, respectively.

6 Conclusions and Outlook

970 This paper aims to ~~achieve a parameterization model for the wind evolution model to predict the wind evolution according to the wind field conditions. Because the concept of the investigate the potential of Gaussian process regression (GPR) in the parameterization of wind evolution. This research has been motivated by the need of lidar-assisted wind turbine control needs the aid of an accurate model for the prediction of the wind evolution~~for accurate models to predict wind evolution, in order to avoid harmful and unnecessary control ~~action.~~actions. In addition, the commonly used 3-dimensional stochastic wind field ~~simulation method can be extended to 4-dimensional by integrating wind evolution, to provide a more realistic simulation environment for this control concept.~~

980 For this purpose, a two-parameter wind evolution model suggested in literature was applied to model the wind evolution. In this research, data from two nacelle-mounted lidars in both onshore and offshore locations were used to estimate wind evolution. Some data of two nacelled-based lidars and an ultrasonic anemometer of two measurements carried out in an onshore and an offshore location were involved in the study. The wind evolution, i. e. the coherence between two measuring points, was estimated using the lidar data and the wind evolution model parameters a and b were determined by fitting the estimated coherence to the ~~The estimated~~ wind evolution was fitted to a two-parameter wind evolution model. The relevant wind field condition variables were derived from the data of the lidars and the ultrasonic anemometer.

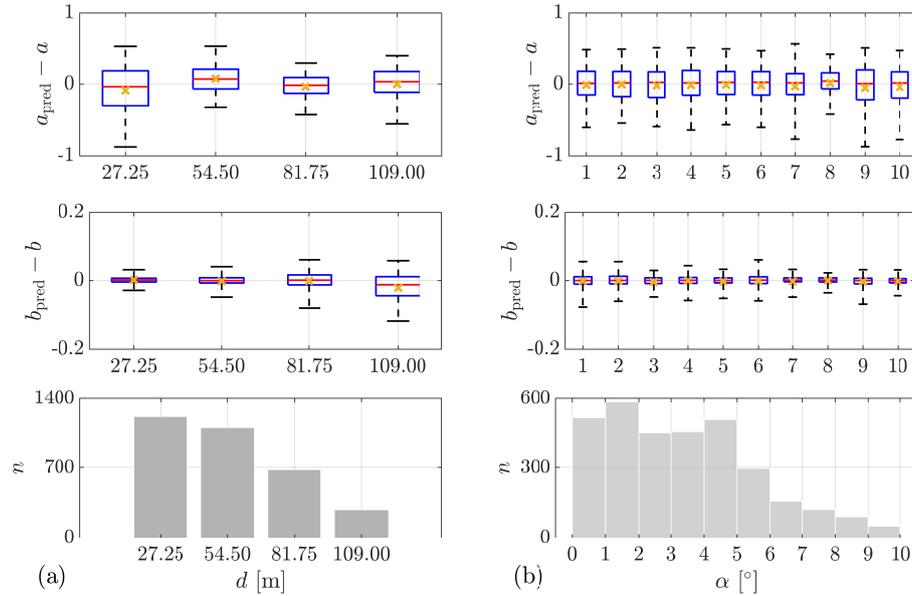


Figure 16. Prediction errors of a (Case 6) and b (Case 11) from LidarComplex with respect to the values of predictors. (a) Measurement separation d . (b) Misalignment angle of wind direction and lidar measurement α . n is sample size. The bottom and top of the boxes indicate the first and the third quartiles, i.e. 25th and 75th percentile, respectively. The lower and upper whiskers show 5th and 95th percentiles. The red line and the yellow cross in the middle indicate the median and mean value, respectively.

Two main outcomes are presented: a statistical analysis of the wind evolution model parameters to reveal, modified from
 985 a model suggested in the literature. To shed light on some characteristics of them and an investigation of the applicability of
 the Gaussian process regression for prediction of wind evolution, a statistical analysis was done for the wind evolution model
 parameters.

In the statistical analysis, the distributions of the wind evolution model parameters of the both measurements show some
 common characteristics, despite different wind field conditions wind-field-related variables and settings of the measurements.
 990 The value ranges of the both wind evolution parameters a (i.e. the decay parameter) and b (i.e. the offset parameter) are very
 similar in the both measurements. The distributions of a and the b seem to follow an inverse Gaussian distribution and a
 Gamma distribution, respectively. The fitted parameters of the probability density functions are different in the both measure-
 ments. It is thus hypothesized We hypothesize that the parameters of the probability density functions should might depend
 on the terrain type. But it was not possible to analyze the influence of the atmospheric stability due to the lack of necessary
 995 data. Moreover, it was observed a strong dependency Moreover, a strong dependence of wind evolution model parameters on
 the measuring separation was observed on measurement separations. The decay parameter a shows a decreasing trend with in-
 creasing measuring measurement separation, while the offset parameter b shows an increasing trend with increasing measuring
 measurement separation.

The applicability of Gaussian process regression to achieve the parameterization model was investigated. The trained models show satisfactory prediction accuracy under a 5-fold cross validation. An investigation was done to explore the potential of using GPR to achieve parameterization models for wind evolution. GPR models were trained with the wind evolution model parameters (i.e. targets) and some wind-field-related variables (i.e. predictors) acquired from the lidars and a meteorological mast. The automatic relevance determination squared exponential kernel was applied to evaluate the relative importance of the predictors for the model and different predictors and to select the essential predictors. It is confirmed that this kernel is capable of selecting the reasonable predictors under different data availability cases. Some interesting insights are given by the predictor selection, e. g. introducing for the models under different data availabilities. The performance of the GPR models was evaluated with the coefficient of determination R^2 and root-mean-squared error (RMSE) using a 5-fold cross-validation. The R^2 of the models in the recommended cases for both targets, under different measurement campaigns and different data availabilities, range from 0.67 to 0.83.

A comparison between the models trained with different predictor combinations provides some interesting insights: 1) GPR models show preference to a fundamental variable than a derived variable when selecting between two related variables. 2) Introducing higher-order wind statistics or (i.e. skewness and kurtosis) as predictors can improve the models. 3) When using travel time as a predictor, the approximation determined with the maximum cross-correlation is slightly preferred than Taylor's translation hypothesis, but the latter could still be an option for the sake of simplification. 4) Introducing one of the predicted targets as predictors can improve the prediction accuracy of the model. The model performance is also evaluated by a visualization of the predicted coherence targets as a predictor for the other can also improve the models, but further research needs to be done to understand the propagation of the uncertainties introduced by the first predicted target. 5) Considering the misalignment angle as a predictor can properly account for its influence on the horizontal coherence. 6) Prediction using sonic data (not measured nearby) does not show any advantages given that it requires many more predictors to exceed the prediction using lidar data.

The predicted coherence is obtained by putting the two predicted parameters into the wind evolution model. To intuitively display how the prediction errors of a and b affect the shape and the position of the predicted coherence in the frequency domain, the predicted coherence and its 95% confidence interval in the frequency domain was visualized for a representative case-study example. The predicted coherence matches the coherence estimated with the measured from data very well, and the 95% confidence interval is relatively narrow. In addition, it was analyzed whether the prediction errors is relevant to the values of the predictors. No obvious relevance can be observed. As a result, it is proven that the Gaussian process regression is capable of achieving a parameterization model with sufficient accuracy for the prediction of RMSE interval was also demonstrated to show the impact of the RMSE of a and b in a more general sense. The RMSE interval turns out to be quite narrow, indicating an overall good model performance. Furthermore, the wind evolution prediction errors of a and b were analyzed with respect to the values of each predictor, shown as boxplots. The results show that, for both a and b , there is no systematic error with respect to predictor values. The prediction of a seems to be less accurate for small travel time and measurement separation. The prediction errors of b show some relevance to the values of standard deviation and skewness of wind speed, travel time, and measurement separation.

1035 ~~For the model~~ There is still space to improve the performance of the parameterization model, ~~there is still space for~~
improvement. Since the performance of any regression model ~~is only possible to be~~ can be only as good as the quality of the
training data, reducing the uncertainty in the training data or increasing the data amount could improve the model performance.
For example, methods to improve the estimation of the coherence and the wind statistics from lidar data are ~~desired~~ desirable.
Moreover, the predictors discussed above do not cover all ~~the~~ possibilities. Introducing new proper predictors could hence also
1040 be easily integrated.

In the future, besides the ideas ~~above-mentioned, it is mentioned above, it would be~~ interesting to involve more measurement
data, especially from different terrain types, to further ~~study if the characteristics of the wind evolution found out in this~~
~~work commonly exist and what are the physical principles~~ investigate whether the wind evolution characteristics found here
occur commonly, and what physical principles stand behind them. Another question ~~needs to be answered is if that needs~~
1045 answering is whether it is possible to achieve a generally applicable parameterization model, and how. Moreover, considering
that the computational time of the model training could be an important issue for some applications, e.g. real-time model
training, it is worth ~~doing a comparison between the Gaussian process regression and~~ comparing GPR with some alternative
algorithms to ~~provide an insight of~~ develop insight into the trade-off between ~~the computational~~ computation time and the
prediction accuracy ~~of the model~~. Furthermore, ~~consider~~ considering the application of the parameterization model using real-
1050 time measurement data as predictors, an additional model will be needed to determine whether the current data meets the
quality requirements to be input into the parameterization model. ~~This will be one of the research focuses in the near future.~~

Last but not least, as mentioned above, the our model concept is very flexible and ~~the its~~ methodology can be applied in
different situations. For example, for other lidar trajectories or even other measurement devices, the model concept can be
modified by replacing the ~~method of the estimation of coherence~~ coherence estimation method. The wind evolution model
1055 and the regression model can also be changed. Basically, one can achieve a parameterization model to meet ~~own specific~~
~~requirements~~ various specific requirements by following the concept and the methodology presented in this paper.

Appendix A

Author contributions. YC conceived the concept, developed the model codes, processed the data, created the figures, conducted the analysis,
and prepared the manuscript. DS and PWC provided general guidance and essential suggestions throughout the process. PWC supervised
1060 the research.

Competing interests. The authors declare that they have no conflict of interest.

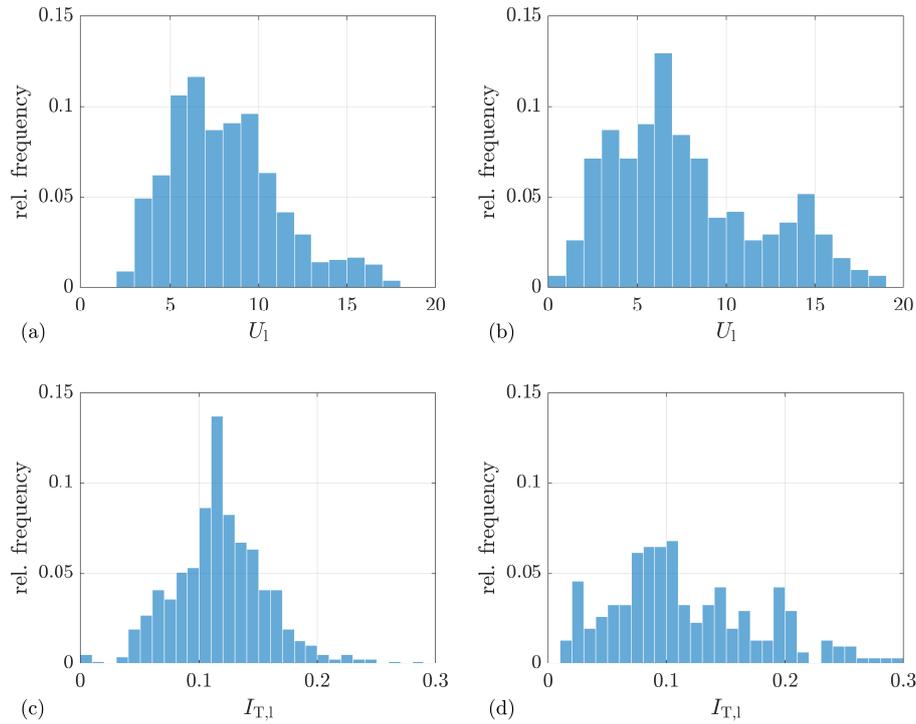


Figure A1. Distribution of lidar measured mean wind speed and turbulence intensity for the selected period. (a) and (c) are from LidarComplex, measurement point at 163.5 m; (b) and (d) are from ParkCast, measurement point at 150 m.

Acknowledgements. This research is carried out within the framework of Joint Graduate Research Training Group Windy Cities supported by the State Ministry of Baden-Wuerttemberg for Sciences, Research and Arts. Some data used in this work is-was generated in the project projects LidarComplex (code number: 0325519A) and ParkCast (code number: 0324330A) which are funded by the German Federal Ministry 1065 for Economic Affairs and Energy (BMW). This publication is supported by the Open Access Publications funding of University of Stuttgart.

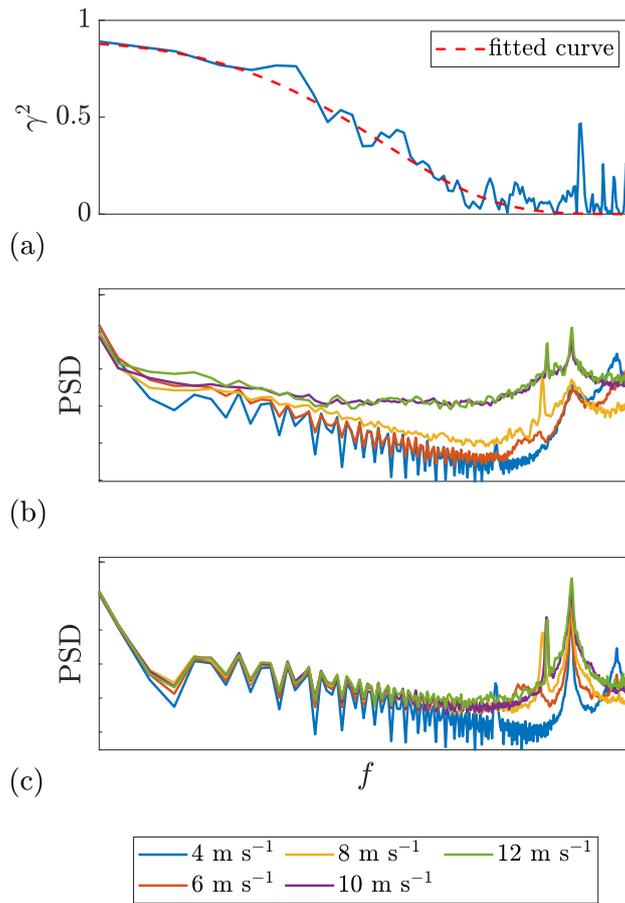


Figure A2. (a) Example coherence: $U_1 = 11.7 \text{ m s}^{-1}$, $d = 81.75 \text{ m}$, $R^2 = 0.95$. (b) and (c) are PSDs of the fore-aft and in-plane tower top acceleration, respectively. The ~~x-axis~~ x-axis is logarithmic. Date and time: 07 Dec. 2013, 12:00 - 12:30. Data Source: LidarComplex. Because of data protection it is not allowed to show any values concerning the turbine properties.

References

- Bossanyi, E.: Un-freezing the turbulence: Application to LiDAR-assisted wind turbine control, IET Renewable Power Generation, 7, 321–329, <https://doi.org/10.1049/iet-rpg.2012.0260>, 2013.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R.: Classification and regression trees, CRC Press, Boca Raton, FL, 1984.
- 1070 Businger, J. A., Wyngaard, J. C., Izumi, Y., and Bradley, E. F.: Flux-Profile Relationships in the Atmospheric Surface Layer, Journal of the Atmospheric Sciences, 28, 181–189, [https://doi.org/10.1175/1520-0469\(1971\)028<0181:FPRITA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2), 1971.
- Cariou, J.-P.: Pulsed Lidars, in: Remote Sensing for Wind Energy, DTU Wind Energy-E-Report-0029(EN), chap. 5, pp. 104–121, 2013.
- Chen, Y.: Parameterization of wind evolution model using lidar measurement, <https://doi.org/10.5281/zenodo.3366119>, 2019.
- Davenport, A. G.: The spectrum of horizontal gustiness near the ground in high winds, Quarterly Journal of the Royal Meteorological Society, 1075 87, 194–211, <https://doi.org/10.1002/qj.49708737208>, 1961.

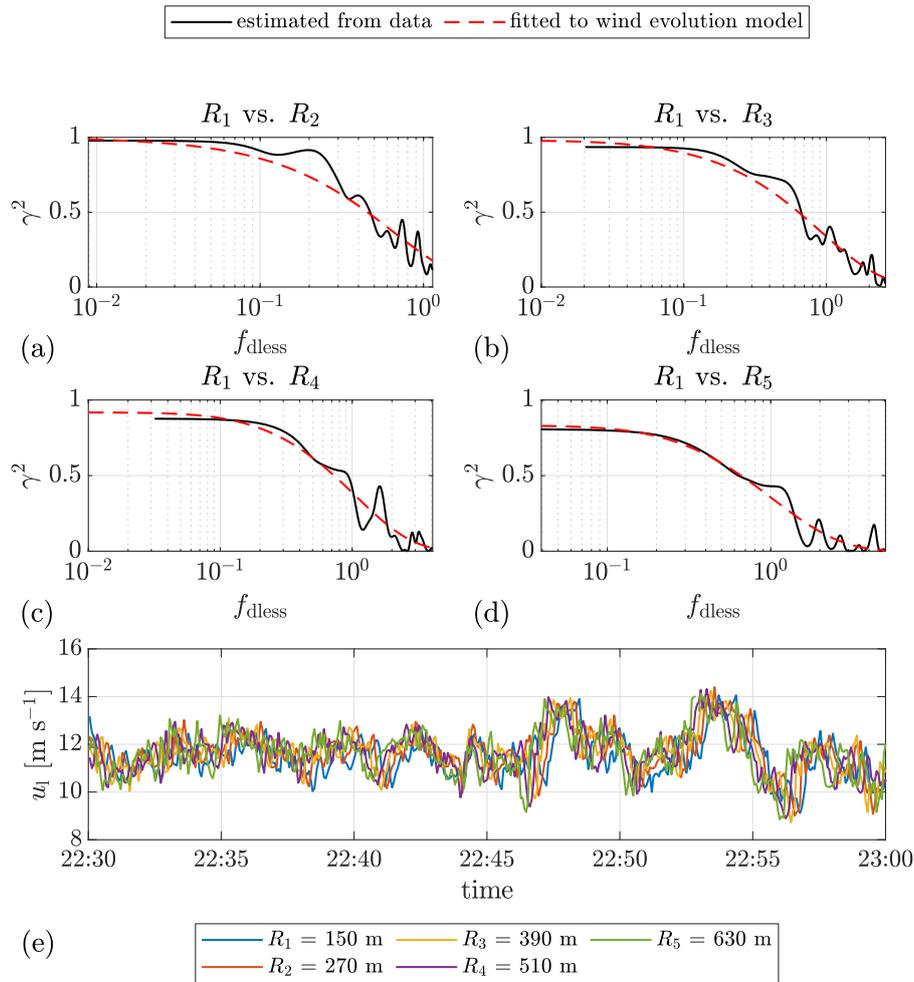


Figure A3. (a)–(d) Example plots of the estimated coherence between the selected range gates R and the corresponding fitted curves. The corresponding measurement distances are 120 m, 240 m, 360 m, and 480 m, respectively. (e) Time series of the lidar wind speed. The mean lidar wind speed is 11.6 m s^{-1} . Date: 12 June 2019. Data source: ParkCast.

Davoust, S. and von Terzi, D.: Analysis of wind coherence in the longitudinal direction using turbine mounted lidar, *Journal of Physics: Conference Series*, 753, 072 005, <https://doi.org/10.1088/1742-6596/753/7/072005>, 2016.

de Maré, M. and Mann, J.: On the Space-Time Structure of Sheared Turbulence, *Boundary-Layer Meteorology*, 160, 453–474, <https://doi.org/10.1007/s10546-016-0143-z>, 2016.

1080 Duvenaud, D.: Automatic model construction with Gaussian processes, *Apollo - university of cambridge repository*, <https://doi.org/10.17863/CAM.14087>, 2014.

Fritsch, F. N. and Carlson, R. E.: Monotone Piecewise Cubic Interpolation, *SIAM Journal on Numerical Analysis*, 17, 238–246, <https://doi.org/10.1137/0717021>, 1980.

Hocking, R. R.: The Analysis and Selection of Variables in Linear Regression, *Biometrics*, 32, 1, <https://doi.org/10.2307/2529336>, 1976.

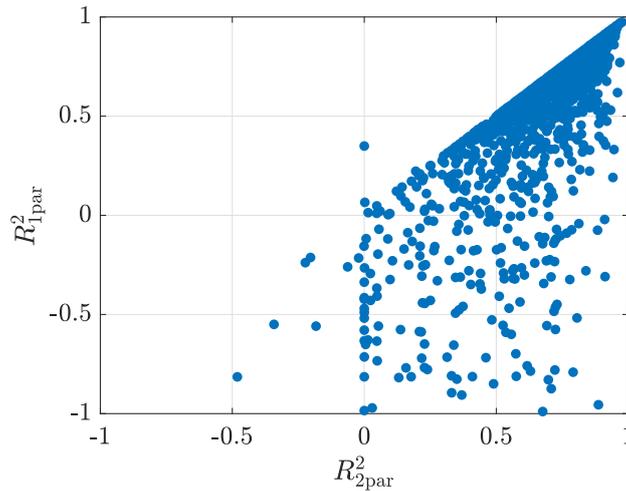


Figure A4. Comparison of R^2 of the fitting to quality (R^2) of the two-parameter wind evolution model (subscription "2par" Eq. (8)) and that of the fitting to the one-parameter one-wind evolution model (subscription "1par" Eq. (5)). The subscripts '1par' and '2par' indicate one-parameter and two-parameter, respectively.

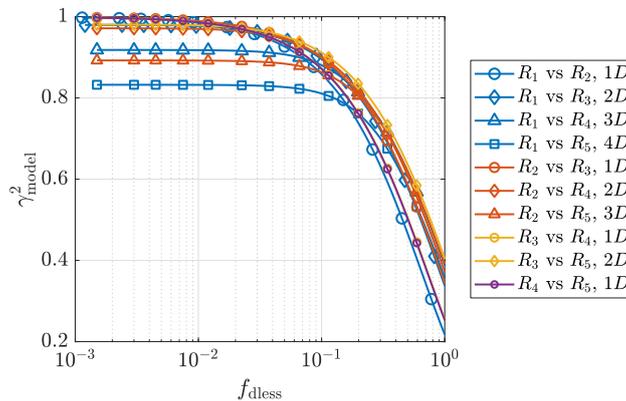


Figure A5. Fitted curves of the estimated coherence between the lidar wind speeds measured at different range gates. The range gate R_1 to R_5 are located at 150 m, 270 m, 390 m, 510 m, and 630 m, respectively. $1D = 120$ m. The mean lidar wind speed $U_l = 11.6 \text{ m s}^{-1}$. Date and time: 12 June 2019, 22:30-23:00. Data source: ParkCast.

1085 J. L. Lumley and H. A. Panofsky: The structure of atmospheric turbulence., Interscience Publishers (Wiley & Sons), New York, 1st edition edn., <https://doi.org/10.1002/qj.49709138926>, 1964.

Joanes, D. N. and Gill, C. A.: Comparing measures of sample skewness and kurtosis, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47, 183–189, <https://doi.org/10.1111/1467-9884.00122>, 1998.

1090 Kelberlau, F. and Mann, J.: Better turbulence spectra from velocity–azimuth display scanning wind lidar, *Atmospheric Measurement Techniques*, 12, 1871–1888, <https://doi.org/10.5194/amt-12-1871-2019>, 2019.

- Kreyszig, E.: Advanced engineering mathematics, Wiley, New York and Chichester, 4th ed. edn., 1979.
- Kristensen, L.: On longitudinal spectral coherence, *Boundary-Layer Meteorology*, 16, 145–153, <https://doi.org/10.1007/BF02350508>, 1979.
- Laks, J., Simley, E., and Pao, L.: A spectral model for evaluating the effect of wind evolution on wind turbine preview control, in: 2013 American Control Conference, pp. 3673–3679, IEEE, <https://doi.org/10.1109/ACC.2013.6580400>, 2013.
- 1095 Lenschow, D. H., Mann, J., and Kristensen, L.: How Long Is Long Enough When Measuring Fluxes and Other Turbulence Statistics?, *Journal of Atmospheric and Oceanic Technology*, 11, 661–673, [https://doi.org/10.1175/1520-0426\(1994\)011%3C0661:HLILEW%3E2.0.CO;2](https://doi.org/10.1175/1520-0426(1994)011%3C0661:HLILEW%3E2.0.CO;2), 1994.
- Levenberg, K.: A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics*, 2, 164–168, <https://doi.org/10.1090/qam/10666>, 1944.
- 1100 Liu, Z., Barlow, J. F., Chan, P.-W., Fung, J. C. H., Li, Y., Ren, C., Mak, H. W. L., and Ng, E.: A Review of Progress and Applications of Pulsed Doppler Wind LiDARs, *Remote Sensing*, 11, 2522, <https://doi.org/10.3390/rs11212522>, 2019.
- Mann, J.: The spatial structure of neutral atmospheric surface-layer turbulence, *Journal of Fluid Mechanics*, 273, 141, <https://doi.org/10.1017/S0022112094001886>, 1994.
- Mann, J., Cariou, J.-P. C., Parmentier, R. M., Wagner, R., Lindelöw, P., Sjöholm, M., and Enevoldsen, K.: Comparison of 3D
 1105 turbulence measurements using three staring wind lidars and a sonic anemometer, *Meteorologische Zeitschrift*, 18, 135–140, <https://doi.org/10.1127/0941-2948/2009/0370>, 2009.
- Marquardt, D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11, 431–441, <https://doi.org/10.1137/0111030>, 1963.
- Mendenhall, W. and Sincich, T.: Statistics for engineering and the sciences, Pearson Prentice Hall, Upper Saddle River, N.J., 5th ed. edn.,
 1110 2007.
- Moré, J. J.: The Levenberg-Marquardt Algorithm: Implementation and Theory., Watson G.A. (eds) *Numerical Analysis*, 630, 105—116, <https://doi.org/10.1007/BFb0067700>, 1978.
- Obukhov, A. M.: Turbulence in an atmosphere with a non-uniform temperature, *Boundary-Layer Meteorology*, 2, 7–29, <https://doi.org/10.1007/BF00718085>, 1971.
- 1115 Oppenheim, A. V., Willsky, A. S., and Nawab, S. H.: Signals & systems, Prentice-Hall signal processing series, Prentice Hall and London : Prentice-Hall International, Upper Saddle River, 2nd ed. edn., 1997.
- Panofsky, H. A. and McCormick, R. A.: Properties of spectra of atmospheric turbulence at 100 metres, *Quarterly Journal of the Royal Meteorological Society*, 80, 546–564, <https://doi.org/10.1002/qj.49708034604>, <http://onlinelibrary.wiley.com/doi/10.1002/qj.49708034604/pdf>, 1954.
- 1120 Panofsky, H. A. and Mizuno, T.: Horizontal coherence and Pasquill's beta, *Boundary-Layer Meteorology*, 9, 247–256, <https://doi.org/10.1007/BF00230769>, 1975.
- Panofsky, H. A., Thomson, D. W., Sullivan, D. A., and Moravek, D. E.: Two-point velocity statistics over Lake Ontario, *Boundary-Layer Meteorology*, 7, 309–321, <https://doi.org/10.1007/BF00240834>, 1974.
- Peña, A., Hasager, C., Lange, J., Anger, J., Badger, M., Bingöl, F., Bischoff, O., Cariou, J.-P., Dunne, F., Emeis, S., Harris, M., Hofsäuss, M.,
 1125 Karagali, I., Laks, J., Larsen, S., Mann, J., Mikkelsen, T., Pao, L., Pitter, M., Rettenmeier, A., Sathe, A., Scanzani, F., Schlipf, D., Simley, E., Slinger, C., Wagner, R., and Würth, I.: Remote Sensing for Wind Energy, no. 0029(EN) in DTU Wind Energy E, DTU Wind Energy, Denmark, 2013.

- Pielke, R. A. and Panofsky, H. A.: Turbulence characteristics along several towers, *Boundary-Layer Meteorology*, 1, 115–130, <https://doi.org/10.1007/BF00185733>, 1970.
- 1130 Pope, S. B.: *Turbulent flows*, Cambridge University Press, Cambridge and New York, 2000.
- Rasmussen, C. E. and Williams, C. K. I.: *Gaussian processes for machine learning*, Adaptive computation and machine learning, MIT, Cambridge, Mass. and London, 2006.
- Ropelewski, C. F., Tennekes, H., and Panofsky, H. A.: Horizontal coherence of wind fluctuations, *Boundary-Layer Meteorology*, 5, 353–363, <https://doi.org/10.1007/BF00155243>, 1973.
- 1135 Sathe, A. and Mann, J.: Measurement of turbulence spectra using scanning pulsed wind lidars, *Journal of Geophysical Research: Atmospheres*, 117, n/a–n/a, <https://doi.org/10.1029/2011JD016786>, 2012.
- Schlipf, D.: *Lidar-assisted control concepts for wind turbines*, Dissertation, 2015.
- Schlipf, D., Trabucchi, D., Bischoff, O., Hofsäß, M., Mann, J., Mikkelsen, T., Rettenmeier, A., Trujillo, J. J., and Kühn, M.: Testing of frozen turbulence hypothesis for wind turbine applications with a scanning LIDAR system, <https://doi.org/10.18419/opus-3915>.
- 1140 Schlipf, D., Haizmann, F., Cosack, N., Siebers, T., and Cheng, P. W.: Detection of Wind Evolution and Lidar Trajectory Optimization for Lidar-Assisted Wind Turbine Control, *Meteorologische Zeitschrift*, 24, 565–579, <https://doi.org/10.1127/metz/2015/0634>, 2015.
- Shannon, C. E.: Communication in the Presence of Noise, *Proceedings of the IRE*, 37, 10–21, <https://doi.org/10.1109/JRPROC.1949.232969>, 1949.
- Simley, E. and Pao, L. Y.: A longitudinal spatial coherence model for wind evolution based on large-eddy simulation, in: 2015 American Control Conference (ACC), pp. 3708–3714, IEEE, <https://doi.org/10.1109/ACC.2015.7171906>, 2015a.
- 1145 Simley, E., Fürst, H., Haizmann, F., and Schlipf, D.: Optimizing Lidars for Wind Turbine Control Applications—Results from the IEA Wind Task 32 Workshop, *Remote Sensing*, 10, 863, <https://doi.org/10.3390/rs10060863>, 2018.
- Simley, E. J. and Pao, L. Y.: *Wind Speed Preview Measurement and Estimation for Feedforward Control of Wind Turbines*, ProQuest Dissertations & Theses, Ann Arbor, 2015b.
- 1150 Taylor, G. I.: The Spectrum of Turbulence, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 164, 476–490, <https://doi.org/10.1098/rspa.1938.0032>, 1938.
- Thresher, R. W., Holley, W. E., Smith, C. E., Jafarey, N., and Lin, S.-R.: Modeling the response of wind turbines to atmospheric turbulence, Tech. Rep. RL0/2227-81/2, Mechanical Engineering Department, Oregon State University, Corvallis, Oregon, 1981.
- Vapnik, V. N.: *The Nature of Statistical Learning Theory*, Springer New York, New York, NY, 1995.
- 1155 Veers, P. S.: *Three-Dimensional Wind Simulation*, 1988.
- Weitkamp, C.: *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*, vol. 102, Springer-Verlag, New York, <https://doi.org/10.1007/b106786>, 2005.
- Willis, G. E. and Deardorff, J. W.: On the use of Taylor’s translation hypothesis for diffusion in the mixed layer, *Quarterly Journal of the Royal Meteorological Society*, 102, 817–822, <https://doi.org/10.1002/qj.49710243411>, 1976.
- 1160 Würth, I., Ellinghaus, S., Wigger, M., Niemeier, M. J., Clifton, A., and Cheng, P. W.: Forecasting wind ramps: Can long-range lidar increase accuracy?, *Journal of Physics: Conference Series*, 1102, 012 013, <https://doi.org/10.1088/1742-6596/1102/1/012013>, 2018.