e-WindLidar: making wind lidar data FAIR

Nikola Vasiljević,Tobias Klaas, Lukas Pauscher, João Correia Lopes,
Daniel Fernandes Gomes, Renato Abreuand and Lars Morten Bardal

IRPWind Report
January 2018

Supported by:

Authors: Nikola Vasiljević, Tobias Klaas, Lukas Pauscher, João Correia Lopes, Daniel Fernandes Gomes, Renato Abreu and Lars Morten Bardal

Title: e-WindLidar: making lidar data FAIR

Parties:
DTU Wind Energy
Fraunhofer IEE
UPORTO
NTNU

Authors contribution: Nikola structured the e-WindLidar project and based on this work work Nikola, Tobias and Lukas done the lidar data conceptualization, which is presented in this report. Nikola wrote Chapter 1, 2, 3, 4, 7, 6.3, 6.4, 8 and Appendix A. João, Daniel and Nikola wrote Chapter 5. Renato wrote Section 6.1, Tobias wrote Section 6.2 and Appendix A.1. Lars wrote Appendix A.2. Nikola, Tobias, João and Daniel contributed in the revision of the document.

IRPWind Report (EN)

2018

DOI:10.5281/zenodo.2478051

Sponsorship:
IRPWind

Contact information:
Nikola Vasiljević:
niva@dtu.dk

Tobias Klaas:
tobias.klaas@iee.fraunhofer.de

Pages: 73
References: 12

Technical University of Denmark
DTU Wind Energy
Risø Campus
Frederiksborgvej 399
DK-4000 Roskilde

www.vindenergi.dtu.dk

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The lidar community is by its nature fragmented and scattered to a large degree. There are many small or bigger groups that operate their instruments and produce data by employing their own best practices. As the result, the data are stored in various formats, and in cases when the data is accompanied with the metadata and data reports, which is rare, no particular standards or uniform templating approach is used to provide them. This has a large negative impact on the data usage and data processing efficiency.

From the perspective of data users, in case when they are willing to use lidar data from different data creators, they would need to get familiar with various storing formats and in majority of cases they will need to be in a constant communication with data creators to properly process data (since data descriptors are usually missing or it is not straightforward to interpret them). Additionally, handling datasets stored in different formats, which usually have different structures, requires development of custom code for data processing. The reusability of such code is limited. Also, in practice, we often encounter that new implementations are performed each time new analysis is underway due to the above-mentioned issues. Overall, the current situation with lidar data represents a large obstacle for the development of the community-based data processing tools, which would substantially improve the data analysis efficiency and generation of new knowledge.

There are numerous reasons for the present situation with lidar data. We can say the same for other types of data collected and used in the wind energy sector and elsewhere. Hence, there is one dominating reason. Until recently the roles of data creators and data users were not separated. Those who created data were the ones who used it and publish results of data interpretation. This led to the situation in which data creators were selecting formats they were accustomed to, while omitting data description or not having a consistent approach in describing data since they were at the same time users of that data. However, the volume of data produced today is simply impossible to be analyzed solely by the research staff who created them. Therefore, the roles of those who create data and those who use data became distinguishing. Despite these facts, the old ways of dealing with data are still preserved.

To change the old habits, and basically have better data (i.e, improve data organization, data description, etc.), an extra effort from data creators is required. Considering that data creators are usually scientific staff, that extra effort in the traditional academic environment is not properly valued since it does not fall under the category of a standard scientific work. In fact, it is usually labeled as an engineering or technical work. However, this extra work is of great importance as it provides the basis for the scientific work that will follow.

Nevertheless, regarding the recognition of this type of work things have evolved recently. Data journals are becoming available (e.g., Scientific Data, Geoscience Data Journal, etc.), where data creators can publish short papers describing their datasets. This provides a possibility for data creators to get their recognition in accordance with the traditional scientific metric, i.e. number of publication/citations in the peer-reviewed journals with impact factor. Also, data repositories, such as B2SHARE of EUDAT, provides a possibility for data creators to publish their datasets and acquire a digital object identifier (DOI), which makes data citable. Particularly related to the wind energy community, this will be made possible throughout the web portal called WindShare. Publishing a short paper in a data journal accompanied by a publication of a dataset in a data repository such as B2SHARE represents an effective approach for data creators to get their recognition. Therefore, mechanisms for acknowledging efforts of data creators have been put in place, which should motivate them to use and/or adopt the best practices for making better data.

A set of guiding principles for making better data, labeled with an acronym FAIR Wilkinson et al. (2016), are gaining momentum in Europe. The acronym FAIR stays for Findable, Accessible, Interoperable and Reusable. How each attribute is related to data is as following:

- Data should be findable
- Humans and machines should be able to gain access to our data (data do not need to be 100% open)
- Data should be interoperable, which means that data should conform to recognized formats and standards
- An additional documentation that further describes data should be provided to support interpretation and reuse of data

Potentially there are two main approaches to make data of one entire sector FAIR. Considering the whole wind energy sector, it is possible to have one large transnational project which deals with integrated activities on making all wind energy data FAIR. The second approach entails many small projects, handled by specialized communities (in our case the lidar community) within the whole sector, which are individually tacking FAIR data issues on one type of data. This approach is often denoted as a divide and conquer. Each approach has it positive and negative sides. The divide and conquer approach is by far simpler to manage, organize and fund.

When applying either one of the two approaches highest success is achieved if the group that is working on the FAIR principles application consists of data creators (i.e., domain experts, those who know their data), IT experts (i.e., those who are familiar with latest advancements in the computer science, databases and such) and early adopters of new data standards (i.e., data users and/or implementers of the FAIR data standards).

For the e-WindLidar project, we selected the divide and conquer approach, and focused our intention to make lidar data FAIR. We assembled a heterogeneous workgroup, which members are lidar and IT experts and early adopters and implementers of the FAIR data standards.

In the following chapters of this report, we will present the result of the e-WindLidar project. The report is organized as follows: Chapter 2 discusses our approach in making data findable, while Chapter 3 is dedicated to a way of making data accessible. In Chapter 4 we present the background of lidar measurement process, which serves as the basis for making data interoperable. Chapter 5 is focused on the description of the lidar data conceptual model. In Chapter 6 we are making a transition from the conceptual model to the actual lidar data format which is based on an existing data format. In Chapter 7 we present a template for reporting data which should improve the reuse of data. Finally, in Chapter 8, we provide our concluding remarks.

Chapter 2

Making lidar data findable

The members of the European Energy Research Alliance, Joint Program on Wind Energy (EERA JPWind), set a goal to create a web-based data search portal, which will collect information on data from cloud distributed data centers, catalog the collected information and provide data users with tools to find data for their needs.

The work package two (WP2) Transfer of knowledge of The Integrated Research Program on Wind Energy (IRPWind), established the basis for the implementation of a data search portal (also known as WindShare Portal). Precisely, the WP2 formulated the information architecture to make data findable. The work carried out within the IRPWind WP2 entailed the selection and extension of a metadata standard for describing the datasets used in the wind energy sector, and the development of wind energy specific taxonomies for a number of elements of the metadata standard Sempreviva et al. (2017).

The Dublin Core (DC) protocol (`http://www.dublincore.org`) was selected for the metadata standard. This standard consists of 15 metadata elements that are used to describe any resource (digital or physical) in the wind energy sector. The elements are intended to be factored out in the form of so-called metadata cards (Figure 2.1). The metadata cards represent a digital counterpart of the paper cards in the library catalogs, which contained information on books. Since 15 elements are general and can be used to describe not only any resource in the wind energy sector but any resource in any domain (e.g., plants, books, data, bolts, etc.), the DC protocol was extended with the additional 7 elements to refine the description of data produced and used in the wind energy sector. For 6 non-DC elements (see Figure 2.1 more information) and two DC elements the IRPWind WP2 defined taxonomies, which contain a specific set of words to tag the data (i.e., domain-specific vocabularies).

The example of a blank metadata card is given in Figure 2.1. It should be pointed out that the metadata card will be provided as a JSON encoded text file based on an extended B2Share schema (`http://e-windlidar.windenergy.dtu.dk/mc_schema.json`). The metadata cards are envisioned to accompany datasets they describe. An example of metadata card for the Perdigão-2015 experiment Vasiljević et al. (2017) can be found here `http://e-windlidar.windenergy.dtu.dk/perdigao2015/mc_Perdigao-2015.json`. They should be either manually uploaded to the future wind energy portal (WindShare, http://it.cener.com/demo/ckan/) or existing metadata catalogs (e.g., https://b2share.eudat.eu). Another approach is to make the metadata cards accessible to web crawlers, which will in turn automatically update web-based metadata catalogs. Once the metadata card becomes uploaded to for example WindShare, this portal will assign DOI to the dataset described by the metadata card. This will make the dataset citable and provides the visibility and acknowledgment to the data creator.

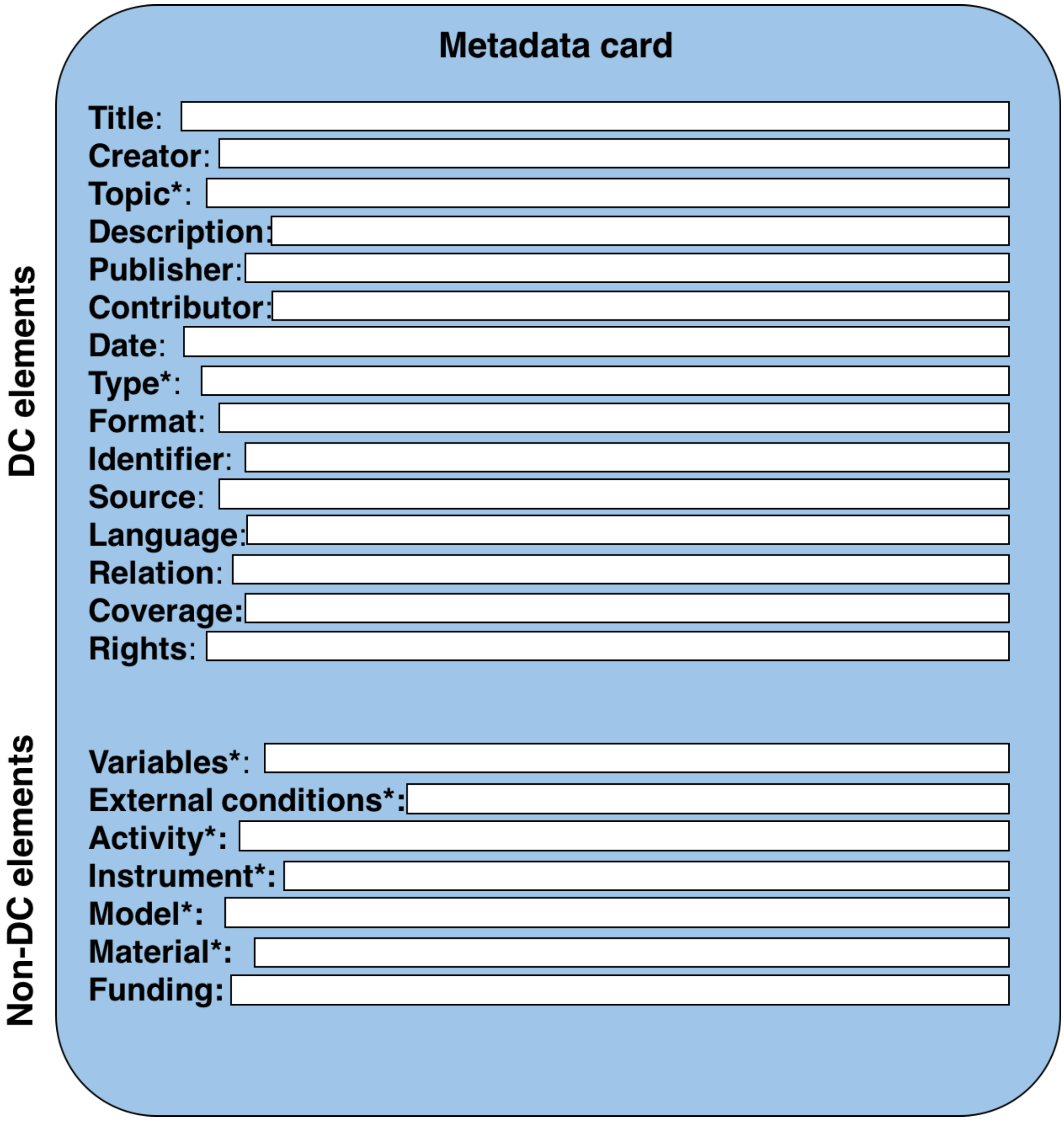


Figure 2.1: The metadata card: elements marked with the asterisk symbol (*) have domain specific vocabularies (i.e., taxonomies)

Chapter 3

Making lidar data accessible

The previous step makes data findable. Also, this step is enabling the data access. The metadata element *Rights* is intended to specify data usage license and also how data users can access the data. In case of the e-WindLidar project, we suggest the use of the CC-BY NC 4.0 license. This license allows data users to perform non-commercial activities with data. They are allowed to distribute, process and build upon the shared data. Data users can perform all these activities as long as they credit the data creators for the original creation. This type of license is recommended for maximum dissemination and use of the data. Potentially, this data usage license can be extended with the requirement for data users to report back how they used data in order to keep the track of the data usage (helps other data users).

In future, data users will have a possibility of a direct and indirect access to the data. The direct access will be provided through the e-WindLidar platform (Figure 3.1), by browsing its data catalog (`e-windlidar.windenergy.dtu.dk`). The prototype of this platform represents a part of the NEWA server/portal (Figure 3.1 and 3.2). For time being, the examples of the FAIR lidar data are available at the e-WindLidar public Git repository: `https://github.com/e-WindLidar`.

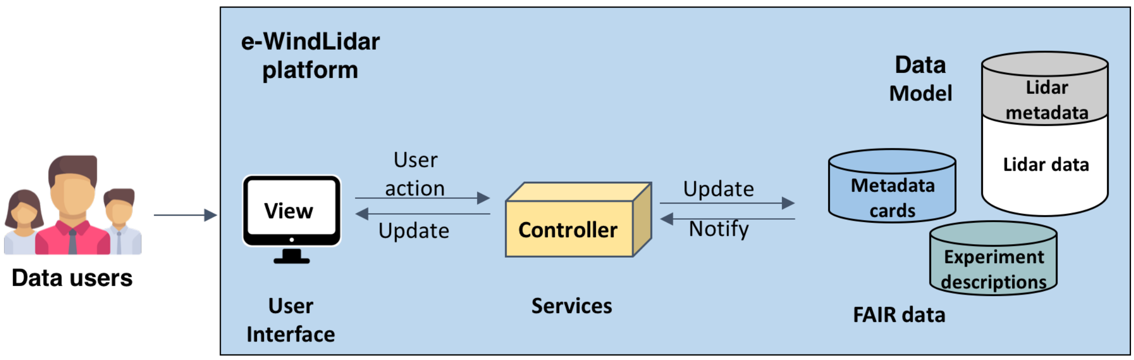


Figure 3.1: The high-level architecture of the e-WindLidar platform

Every dataset that the e-WindLidar platform will offer to data users will be accompanied by a metadata card. The metadata cards will be exposed to the WindShare's web crawler (Figure 3.2), which role is to constantly update WindShare's metadata catalog. As such, through WindShare data users will have an indirect access to data. It is important to state that for datasets which metadata cards will be exposed to the web crawler the WindShare portal will DOI, thus making data citable.

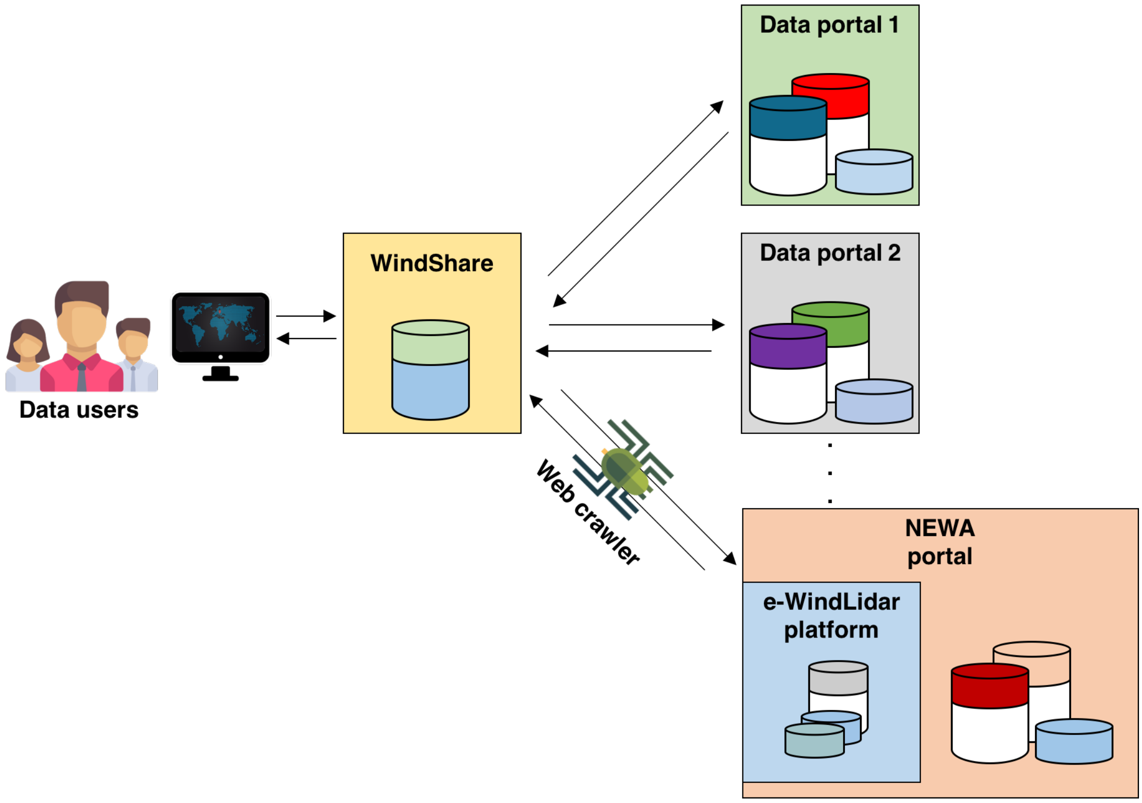


Figure 3.2: Data discovery through the WindShare portal

Chapter 4

Understanding lidar data

To select and/or modify an appropriate standard for storing the lidar data it is essential to understand well the background of the lidar measurement process and what the resulting lidar data represent. We will start with a detailed description of the lidar measurement process while outlining a set of lidar data descriptors. Following this step, we will proceed with a categorization of data products that can be derived from the lidar measurements. Later on in this report for one of the data products we will build a UML representation. Finally, based on the previous results we will select a data standard and adapt it to suit the needs for storage of lidar measurements.

4.1 Description of the lidar measurement process

Lidars acquire wind observations remotely, without contact with the moving air. They do this by emitting the laser light and detecting the Doppler shift in the light backscattered by the aerosols particles (Figure 4.1). The particles are small enough to be advected by the wind (ref Tatarski and Huffaker). The Doppler shift, the frequency difference between the emitted and backscattered light, is a direct measure of the radial or Line Of Sight (LOS) velocity. This velocity is equal to the aerosols particles velocity (i.e., wind velocity) projected on the laser light propagation path.

The emitted light can be represented as a pulse train or a continuous wave (CW). Accordingly, there are two lidar technologies used to sense the wind (i.e., pulsed and CW technology). Each technology has its advantages and disadvantages.

CW lidars can provide high-frequency measurements (up to 400 Hz) of the flow while probing the atmosphere with a relatively small probe length (down to a few cm) which allows resolving small flow features in high resolution. Since the LOS speed at a given point is resolved by focusing the laser beam, the probe length increases with range. This limits the maximum range to about 150 m. Furthermore, CW lidars can only focus light at a single point at a time which means that they can only simultaneously resolve wind speed from one location in the atmosphere. Therefore, the focus distance determines the range where the measurements are

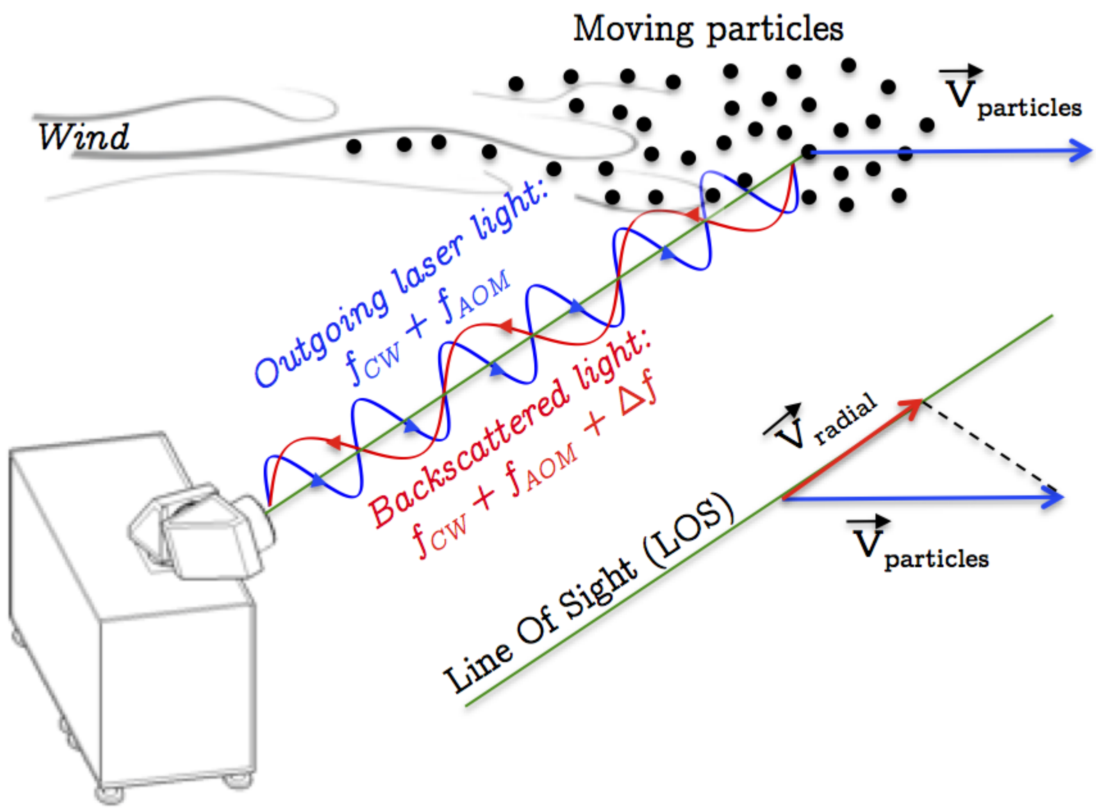


Figure 4.1: Lidar measurement process: f_{CW} (frequency of the laser source), f_{AOM} (carrier frequency) and $\Delta\ f$ induced frequency difference (Doppler shift) due to the motion of particles. (figure source: Vasiljević (2014))

taken. Overall, CW lidars are ideal for the measurements of turbulent flow features within a relatively small area of the atmosphere.

Pulsed lidars have a larger probe length (minimum 25 m) and lower measurement frequency (10 Hz at best, typically 1 Hz) than CW lidars. Since they emit laser pulses, which length is fixed over time, the resulting probe length is constant with range (typically about 30 m). Furthermore, pulsed lidars can simultaneously retrieve radial velocity from a number of ranges along the laser light propagation path. This number is limited by the computational power of the lidar. This particular characteristic of ranging compensates for a lower measurement frequency since at any given measurement rate pulsed lidars can provide a "snapshot" of the atmosphere up to several kilometers along a single LOS. In case of the pulsed lidars the range determination is done by measuring the time of flight of the laser pulse. Depending on their construction and atmospheric conditions, pulsed lidars can achieve maximum range of up to 30 km Kameyama et al. (2012). Due to their characteristics, pulsed lidars are ideal for measurements of mean flow fields within a large area of the atmosphere (e.g., Berg et al. (2015)). Generally speaking, a hybrid lidar system consisting of both pulsed and CW lidars is ideal for obtaining a detailed picture of the atmospheric flows Vasiljević et al. (2017).

As we can see the difference in the technology of producing the laser beam has several implications on the wind speed measurements, being that the most important are: (1) measurement frequency, (2) maximum range, (3) range determination, (4) number of simultaneous radial velocity measurements and (5) probe length characteristic.

Additionally, there is a slight difference in the way how pulsed and CW lidars extract the radial velocity from the light reflected by the moving aerosols. In principle, any lidar directly measures the backscatter signal, which is given as the output of the photodetector (i.e., photocurrent), while the radial velocity is extracted by analyzing this signal. The backscatter signal can be represented as a sine wave, where the wave frequency carries the information on the Doppler shift and thus radial velocity. To extract this information, the first step is to sample (i.e., digitize) the backscatter signal, which results in some hundreds of sample points. The usual sampling frequency is the range between 125 to 250 MHz.

In case of the pulsed lidars, because they measure multiple ranges simultaneously, the entire set of sampled points is split into several subsets. Each subset of points corresponds to a different range at which the radial velocity is to be determined. Afterwards, on each subset, the Fast-Fourier-Transform (FFT) is applied and a number of Doppler spectra is produced from a single pulse propagation through the atmosphere. Since the CW lidars sense the wind speed from one single range, all samples are used to produce a single Doppler spectrum.

For several reasons, such as the elimination of the noise and improvement of the signal-to-noise ratio (insert references), Doppler spectra from several backscattered signals are accumulated and averaged. Afterwards, the averaged Doppler spectra are analyzed to extract the radial velocity. The extraction of the radial velocity in case of the pulsed lidars is often done by applying a maximum-likelihood estimator (Sobolev and Timokhin, 2014) on the averaged spectra, while on the other hand for the CW lidars a centroid method is usually used for this purpose Courtney et al. (2008).

Lidars are either configured to measure in staring (one fixed beam direction), step-stare (several fixed and discrete beam directions) or scanning/sweeping (beam direction continuously varies) modes. In the staring mode, the beam direction is fixed with respect to the beam origin. In the step-stare mode, the beam is steered over several discrete directions, where for each direction the Doppler spectra are accumulated and processed. During the Doppler spectra accumulation, there is no change in the beam direction with respect to the beam origin. In the last mode, the beam is configured to scan through the atmosphere. In this case, the Doppler spectra are accumulated not from a single beam direction. In any of these cases, the beam origin might be fixed (a stationary installation of lidar) or flexible (a mobile installation of lidar such as a lidar on a buoy).

The radial velocity alone can be used in data analysis (e.g., to estimate wind turbine wake geometry, ref). Usually, this is done for qualitative analysis of the flow phenomena (e.g, gravity waves Rodrigues et al., 2016). However, typically radial velocity measurements are converted to the wind vectors. In case of the single lidar operation, the radial velocity measurements from several beam directions are combined to reconstruct the wind vector (Figure 4.2). This requires so-called diverging scanning strategies (e.g., VAD and DBS) and application of single-Doppler wind vector retrievals (ref ref). Another approach is to use several independent lidars (two or three) and configure them to intersect their laser beams at one or several points in the atmosphere depending on the goal of the measurements (Figure 4.2). In the case of multiple-lidar configuration so-called converging scanning strategies are employed, while depending on a number of lidars used dual- or triple- Doppler retrievals are applied to reconstruct the wind vectors from the resulting radial wind speed measurements.

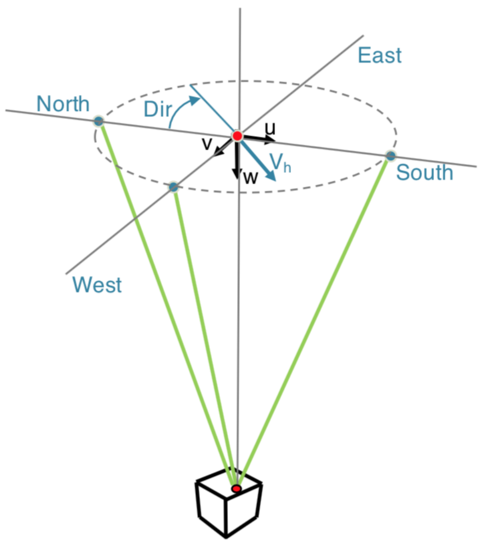


Figure 4.2: Single and multi lidar measurements of the wind vector in a single point.

4.2 Lidar data descriptors

The lidar measurement process is rather delicate and complex. This means that a particular consideration has to be given to the selection of information that needs to be stored in order to better comprehend and use data. It is essential to provide a number of data descriptors aid the data usage. We selected a number of descriptors that are organized in the following categories: (1) general information about the dataset, (2) information about the lidar that was used to obtain measurements, (3) measurement configuration of the lidar during the measurements and (4) data provenance.

The first category of the lidar data descriptors provides a general information about the recorded dataset. The lidar data descriptors of the first category are basically the elements of the metadata card (Figure 1) extended with the following information: (1) convention used to label (recorded) information in the data files, (2) version of the convention, (3) data product (see the following section), and (4) general comments about the dataset.

The second category entails details about the lidar that was used to record data. The descriptors that are a part of this category are: (1) lidar accuracy (LOS speed, pointing and ranging accuracy), (2) lidar technology, (3) type of lidar, (4) lidar installation type, (5) convention on flow direction encoding (i.e., how to interpret the radial velocity sign), (6) lidar labeling (e.g., lidar name, serial number, owner, etc.) and (7) any comment related to the instrument.

The third category indicates details about the configuration of the measurement process. The lidar data descriptors that fall under this category are: (1) retrieval type, (2) linked lidars (in case of multi-Doppler retrievals), (2) probe length, (3) scan strategy type, (4) scan strategy execution type, (5) number of LOS measurements, (6) number of the range gates per LOS measurement, (7) author of the measurement configuration and (8) any comment related to the measurement configuration.

The last category of descriptors holds the information on the data provenance, thus providing the data traceability (what happened to data and when) and visibility of personnel that performs data processing. The descriptors that fall under this category are: (1) date and time when data modification took place, (2) what kind of modification was performed on data, (3) what kind of parameters were used to configure the algorithm that executed the modification, (4) on what data product the modification was executed, (5) what is the resulting data product, (6) who executed the modification, and (7) additional comments related to the data modification.

4.3 Lidar data products

Based on the description of the lidar measurement process one can distinguish a data product scheme that consists of five (major) levels of data products. Among these five products, there is one fundamental data product level and the additional four resulting products. The data products, from the lowest to the highest level, are: (1) Level 0 (L0) product - backscattered signal, (2) Level 1 (L1) product - Doppler spectra, (3) Level 2 (L2) product - radial velocity, (4) Level 3 (L3) - reconstructed wind velocity and (5) Level 4 (L4) - flow characteristics. The data workflow based on the description of the five data products and their sequential interconnection by means of different techniques is given in Figure 4.3.

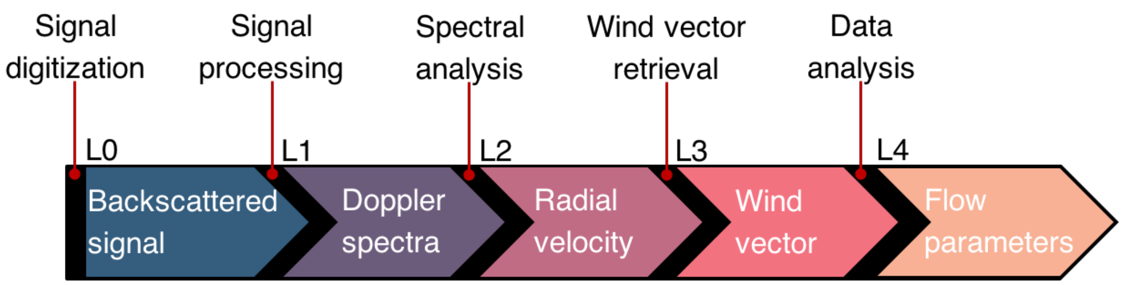


Figure 4.3: Data workflow

The sampled output of the photodiode, which is basically a digitized version of the backscattered signal, represents the fundamental data product or L0. By applying signal processing techniques on L0 data products, Doppler spectra, or L1 data products are produced. From the spectra, using estimation techniques such as the maximum-likelihood estimation, the Doppler shifts, accompanied with carrier-to-noise ratio (CNR) and spectral broadening are estimated, and the resultant LOS wind speeds are calculated. The results of the estimation techniques represent L2 data products. The L2 data products are usually produced in real-time by commercial lidars.

Depending on the number of lidars used in a measurement campaign either single- or multi- Doppler reconstruction techniques are applied to L2 products to retrieved wind vectors. The retrieved wind vectors represent L3 data products. Finally, by analyzing L3 data products, and applying end-user algorithms, diverse parameters of the flow are extracted (e.g. a 10 minute first- and second- order statistics, the wind turbine wake geometry) or data is used to validated flow models and theories. The application of end-user algorithms on L3 data products leads to the creation of L4 data products.

Besides the five major data products, one should expect to have additional sub-level modifications of data at each product level (Figure 4.4). For example, prior conversion of the L2 to L3 products, the L2 products undergo filtering in order to flag erroneous radial velocity estimates which are then excluded in the wind vector retrieval.

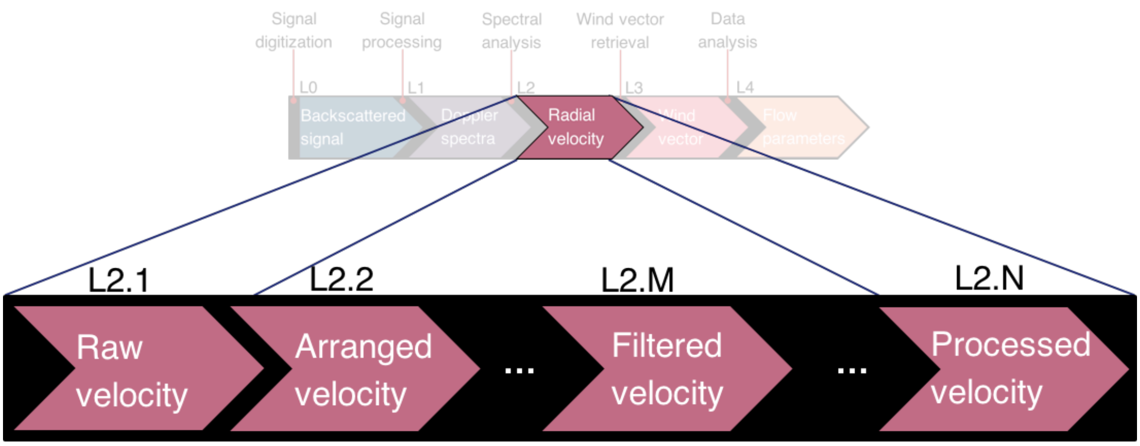


Figure 4.4: Sub-level data workflow

Chapter 5

Lidar data conceptual model

This section presents the conceptual model for L2 lidar data product.

5.1 Introduction

The conceptual model is an abstraction that includes the entity types that represent things or concepts, and relationships between them that are relevant to a specific domain of knowledge, to which we refer as to the Universe of Discourse (UoD). For example, the long-range Windscanner, or the Halo lidar are domain entities of the same entity-type (called Instrument later).

The UoD is characterized by the use of domain-specific vocabularies (e.g., phrases or words) for the communication between parties (stakeholders) involved with L2 lidar data product. The conceptual model serves to:

- Enhance stakeholders understanding of descriptors, variables and dimensions that go to the common data representation proposed by this report.
- Facilitate the communication between stakeholders such as wind energy and atmospheric researchers, lidar developers and software engineers, who develop code to obtain data products or to transform lidar data to the common representation.
- Provide a reference for system designers to extract system specifications.
- Document the system for future reference and provide means for collaboration.

A conceptual model can be described using various notations, but the Unified Modeling Language (UML) has emerged as a typical choice, particularly in Computer Science[1] Rumbaugh et al. (2004).

Using the UML notation, the conceptual model is described by a Class Diagram in which classes represent concepts (the entity-types of the UoD), relationships rep-

[1] For detail on the UML Class diagrams see: Scott W. Ambler, UML 2 Class Diagrams: An Agile Introduction, online at http://www.agilemodeling.com/artifacts/classDiagram.htm

resent relationships between concepts (entities of the UoD) and include the roles taken by the entities that participate in the relationship. Cardinality is represented to qualify the number of the entities participating in the relationship. The class attributes represent the properties of the entities of the domain. The aforementioned will become clearer in the example of lidar data.

In the following sections, the lidar data UoD is divided and explained through several UML class diagrams. Breaking the UoD on several diagrams was done to facilitate reading of this document. Nevertheless, it must be emphasized that the diagrams are the formal description of the included concepts and due to the UML notation also their associated semantic, thus the ambiguity of the (natural language) discourse is minimized. This way, the “phrases” of the UoD are well understood by actual and future stakeholders of the L2 lidar data product[2].

5.2 Observations

The central role of a lidar is to observe the wind at a certain location(s) in the atmosphere (the ”feature of interest”). As we can see from Section 5.1 the key result of this process is the radial velocity, which is estimated from the Doppler spectra.

Figure 5.1 represents the ”feature of interest” concept that is one of the two other concepts: the Doppler spectra, which is a ”feature of interest” of L1, and the Observation, which is a ”feature of interest” of L2[3]. Moreover, each observation (an entity of class Observation) is calculated using several spectra acquired over a certain time interval to which we refer as the accumulation time (see accumulation in Section 5.4).

Besides the estimated radial velocity, additional attributes are extracted from the spectra (Figure 5.2). These parameters are: (1) the carrier-to-noise ratio (CNR), which is used to investigate a goodness of the radial velocity estimation, and (2) spectral width, which is an indicator of the wind speed distribution within the probe length.

The above described three parameters, radial velocity, spectral width (or broadening) and CNR, are observed at a certain location (the beam steering position and distance along LOS) in the atmosphere and over a certain time period. This is modeled in the class Observation of Figure 5.3.

The information on the time recorded when the observation took place (the recorded time) is provided by *one* entity of the class Time. We assume that the GPS driven

[2] A disclaimer must be made here: the diagrams are not complete as they do not include extra UML constraints needed to represent all UoD. Our approach is to give ”just enough” information to the stakeholders.

[3] In the UML representation, the two subclasses are related with the parent class by the so-called IS-A relationship. To be more precise it should be added the restriction that the relationship is complete (one feature of interest must either be one L1 Doppler spectra or a L2 Observation) and disjoint (it can only be one, e.g. they do not overlap).

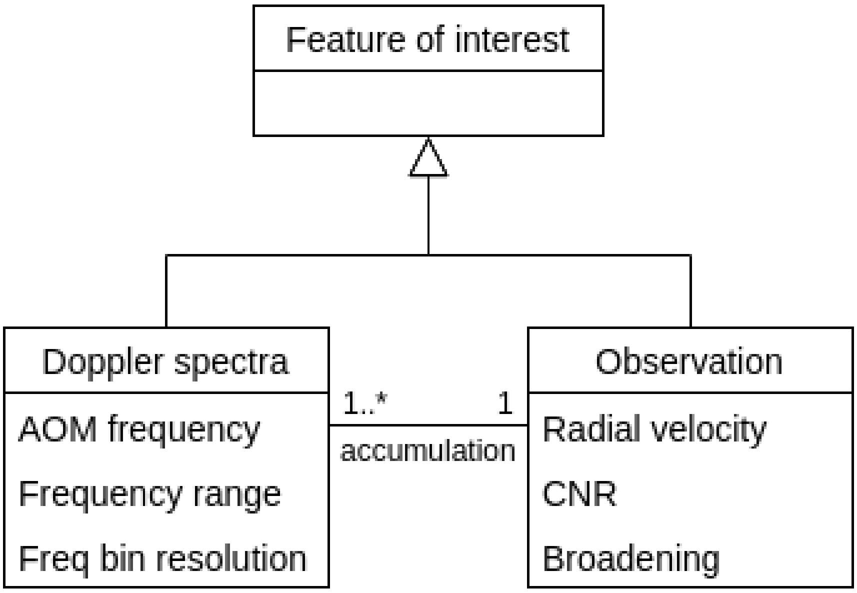


Figure 5.1: UML representation of a Feature of Interest

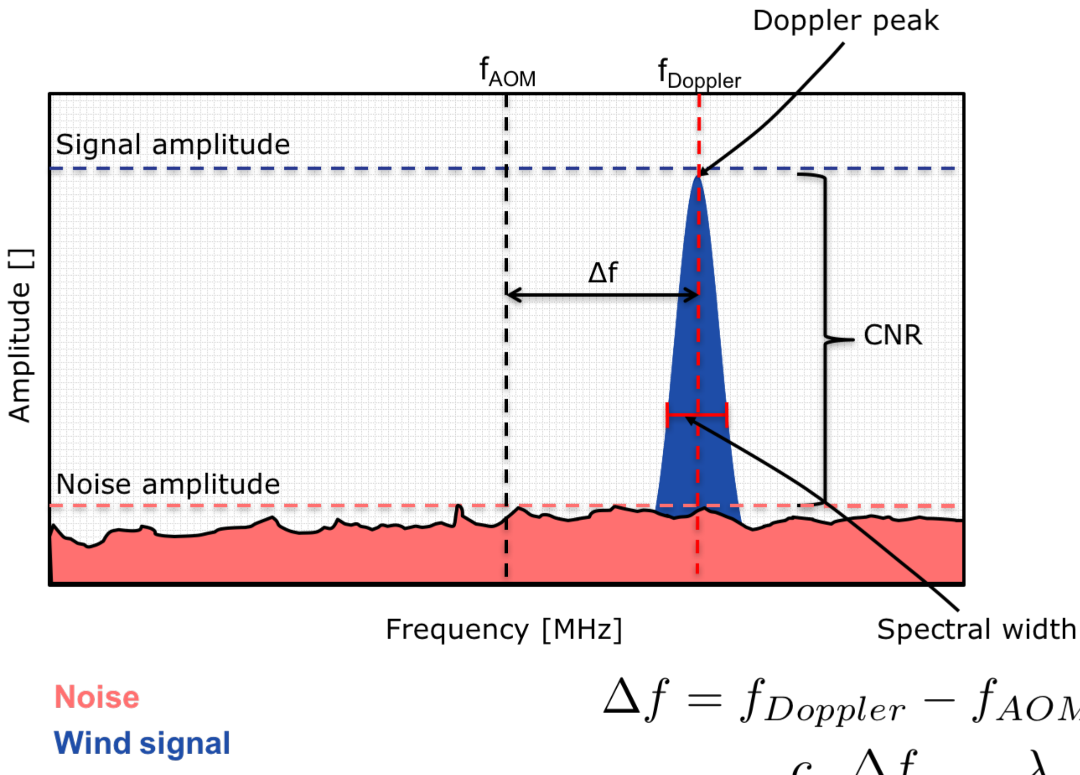
clock is providing the (accurate) *timestamp* information and, thereafter, instances of the the class Time will be used as the representation.

The position of the observed feature of interest (the BeamSteering position) and the beam origin (location plus orientation, called 6DOF) are described in the next section together with the coordinate systems used. The time start describes the GPS time when the observation started the accumulation of Spectra.

5.3 Positions and coordinate systems

The location where the observed parameters are acquired is typically expressed in terms of the coordinates of a spherical coordinate system which origin is collocated with the laser beam origin. We refer to this coordinate system as the beam steering coordinate system (see Figure 5.4). The coordinates of the beam steering coordinates system are: azimuth angle, elevation angle and range as represented by BeamSteering class later in the text. When defining the geometry of a scanning strategy (called configuration later) the same coordinate system is used. Thereafter, the class BeamSteering will be used to represent this coordinate system.

Ideally, in case of the perfect leveling and orientation of the lidar, the azimuth and elevation angles of 0° means that the beam is directed to North parallel with the flat surface. The positive increase in the azimuth angle means that the beam direction is rotated clockwise, thus for the azimuth angle of 180° the beam is directed to South. In case of the elevation angle, the positive increase of this angle means that the beam is rotated counterclockwise, thus for the elevation angle of 90° the beam is fully vertical. The range is simply a distance between the beam origin and the point of interest along the beam. To relate the beam steering coordinate system to a certain geographical coordinate system (e.g., UTM), and thus express the geographical



$$\Delta f = f_{Doppler} - f_{AOM}$$

$$V_{LOS} = \frac{c}{2}\frac{\Delta f}{f_{laser}} = \frac{\lambda}{2}\Delta f$$

Figure 5.2: Doppler spectra: f_{laser} - frequency of the laser source, f_{AOM} - carrier frequency, Δf - Doppler shift, V_{LOS} - LOS speed, c - speed of light and λ - wavelength

location of observation, (ideally) a user needs to know only the position of the beam origin in that coordinate system and a coordinate transformation function.

In practice, during the lidar installation perfect leveling and orientation of the lidar is almost never achieved or simply it is not possible (e.g., lidar installed on an operational wind turbine). This results in the rotation of the beam steering coordinate system which corresponds to the rotation of the lidar with respect to the ideal case: pitch, roll and yaw not zero.

Furthermore, for non-stationary lidar deployments besides the deviation of the lidar leveling and orientation from ideal values also the lidar position, and thus the beam origin position, is not fixed (to be modeled later). This results in the rotation and translation of the beam steering coordinate system. The rotation of the beam steering coordinate system, as mentioned before, corresponds to the rotation of the lidar, while the translation corresponds to the translation of the beam origin with the respect to its initial position.

For many practical reasons, the actual location of measurements should be expressed in terms of the coordinates of the beam origin coordinate system considering the ideal case (i.e., a stationary perfectly leveled and oriented lidar) plus deviations of the lidar six degrees of freedom (6DOF): the beam origin position, and the beam origin orientation (yaw, pitch and roll).

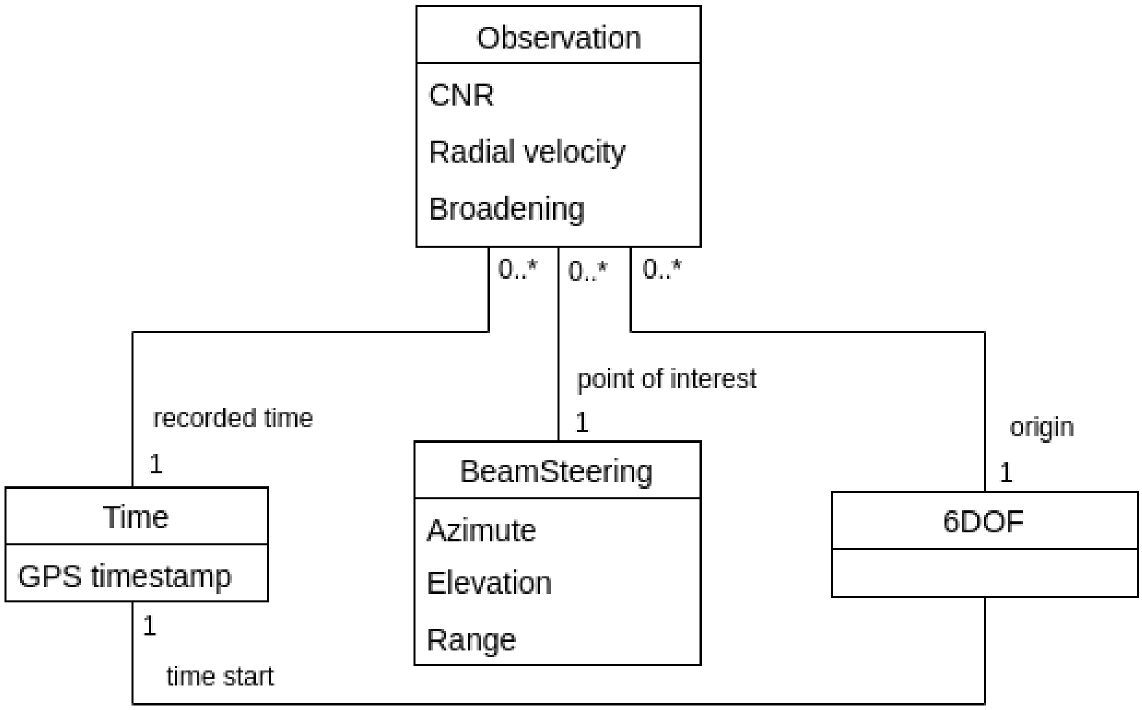


Figure 5.3: UML representation of a single lidar observation (i.e., one measurement point)

The origin of the beam (the initial deviation) is represented by one entity of the the class 6DOF, that relates to one instance of Time (t start of the accumulation), one Position (the beam origin), and the beam Orientation (yaw, pitch and roll), as represented in Figure 5.5. The recorded time is in the middle of the accumulation time used as explained in the next section.

5.4 Accumulation time

As mentioned before, the radial velocity, CNR and spectral broadening are estimated from the accumulation of Doppler spectra. In the most complex scenario, during the spectra accumulation, both the beam and lidar can be non-stationary. For this reason, in Figure 5.6 instances of Observation are related with several deviations (instances of 6DOF) and each deviation is related with one Time. As before, the deviation includes the 6DOF: a position (coordinates) plus orientation (yaw, pitch and roll).

As all the concepts were already introduced (represented by the box with only the name of the UML class) only the relationship between Observation has to be changed to a composition (represented as black diamond) to denote that one observation is composed of a set of deviations (entities of 6DOF) in the time interval: t-start to t-end (see Figure 6.6).

5.5 Installation type

As mentioned before, the lidar installation type has a large impact on the lidar observations.

In some cases, the lidar will have fixed position and orientation during measurement campaigns. However, in other deployments, the lidar is installed on a mobile platform, such as a buoy, which results that the 6DOF of lidar will change over time. In other words, there is a trajectory of the lidar beam origin, which incorporate translation and rotation of the beam origin, connected to all observations acquired by the lidar.

The class Installation in Figure 5.7 has two subclasses, which are Mobile and Fixed to describe the two potential scenarios of the lidar installation. One installation is associated with one instrument and one configuration, which are described elsewhere.

Mobile installations have impact on the observations, and deviations (see Figure 5.7), as the non-stationary installation of the lidar will (continuously) change 6DOF of lidars, which will have impact on the location (relationship between one deviation and one coordinate) where the observations are taken, but also on the observations themselves. Due to the motion of the lidar, the estimated velocity from the Doppler spectra represents the summation of the velocity contributions from the moving aerosols and the lidar motion itself.

For the mobile installation, there is a trajectory while for a fixed installation there is one lidar origin.

5.6 Configurations

So far we have been explaining a single observation in the atmosphere. In fact, lidars are always configured to perform observations at a number of locations in the atmosphere.

Each configuration is identified by a scan id and it is defined by probe length and number of accumulations to be used in the LOS estimation (see Figure 5.8). The configurations have an author, that is affiliated to an Institution, and are related to an instrument (a lidar) during a certain time interval (from date start to date end), as it is represented in by Configuration in Figure 5.8.

We can see that either the configuration of lidar is automatic or it is made by one or several authors, who might be affiliated with one or several institutions. The configuration is used to set a lidar to perform measurements over a certain time period. In the model, this is reflected by means of the attributes date start and date end.

There are several different configuration types represented in Figure 5.9. The lidar could be configured to measure from a single point along the laser light propagation path (i.e. the LOS), to which we typically refer as a range-gate (represented by the Range-Gate configuration), or from several range-gates along the same LOS (represented by LOS) which is a set of points (represented as aggregation), and from a set of LOS (i.e. Multi-LOS). In the multi-LOS configuration the LOS measurements can be done by means of well-established scanning strategies (PPI, RHI, DBS, etc.) or by means of more complex and sophisticated scanning strategies[4] (usually denoted by Complex).

During the collection of the LOS measurements, as discussed earlier, the beam can be fixed or it can traverse. To describe this in Figure 5.9 the class Process is used, together with its two subclasses Sweep and Step to indicate how the LOS measurements are collected (the process).

5.7 Instruments and calibrations

The concepts related to one Instrument (the lidar), namely the accuracy (the calibration), labeling, classification, ownership, operation and calibrations are described in Figure 5.10.

The class Calibration represents the instrument calibration, with the attributes measured and a relationship with the institution that made it. The calibration is made for zero or more instruments and, for each calibration, there one associated date (an instance of Time).

The two subclasses of Instrument represent the technology of the lidar: the class Pulse for pulse lidars, and the class CW to represent the Continuous Wave lidars.

Instruments have several attributes: serial number, product name, special name, range calibration and each instrument belongs to with one or more Institutions (entities of Institution) and is operated by one or more Institution, as well. Furthermore, each instrument may be installed according to the installations types represented in Figure 5.11: ground-based, floating, nacelle-based, blade-based, spinner.

[4] The scanning strategy used in Perdigao 2015 is an example of such configurations: https://www.youtube.com/watch?v=CvnJAev69J0

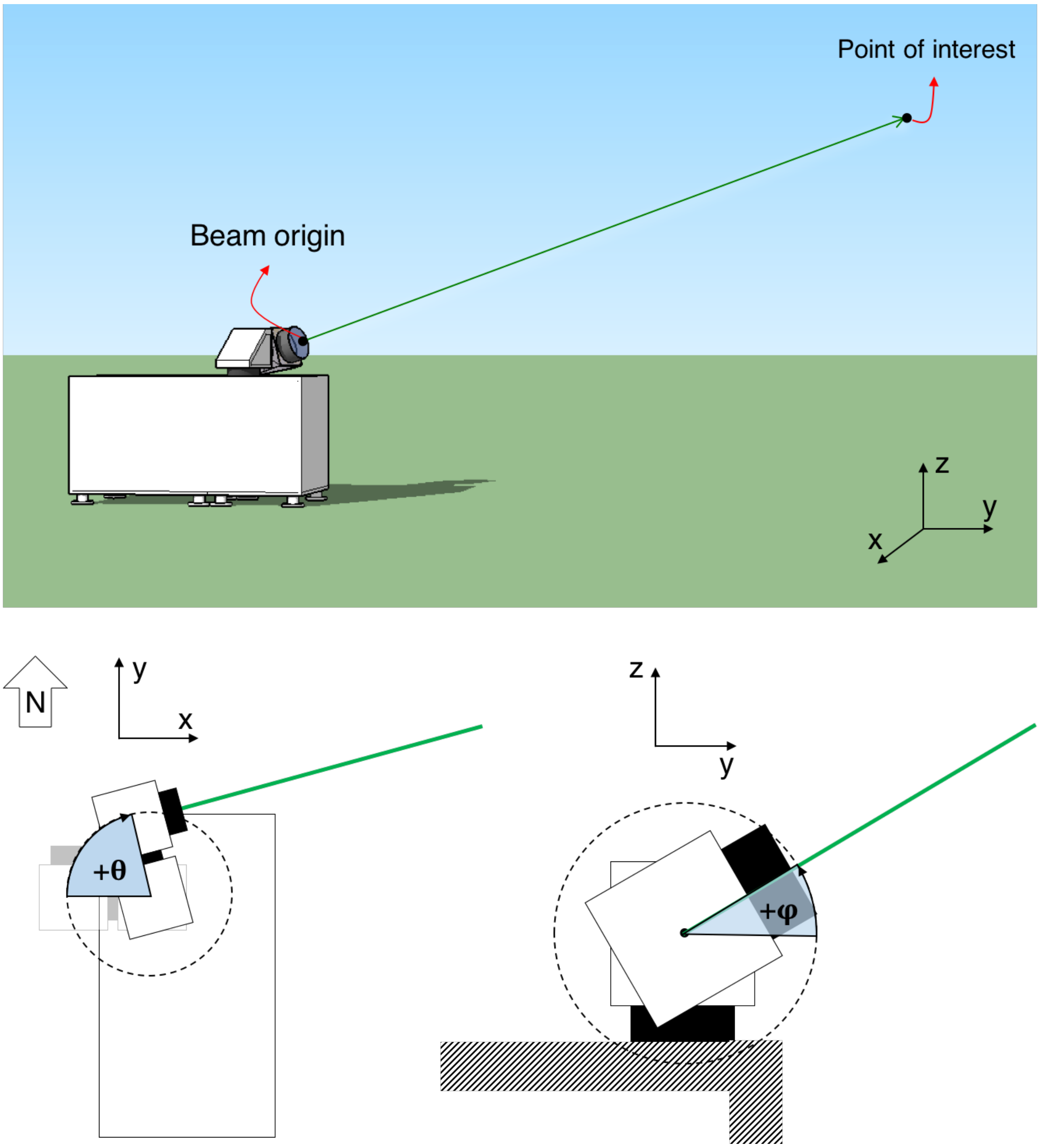


Figure 5.4: Beam steering coordinate system: θ - azimuth and φ - elevation

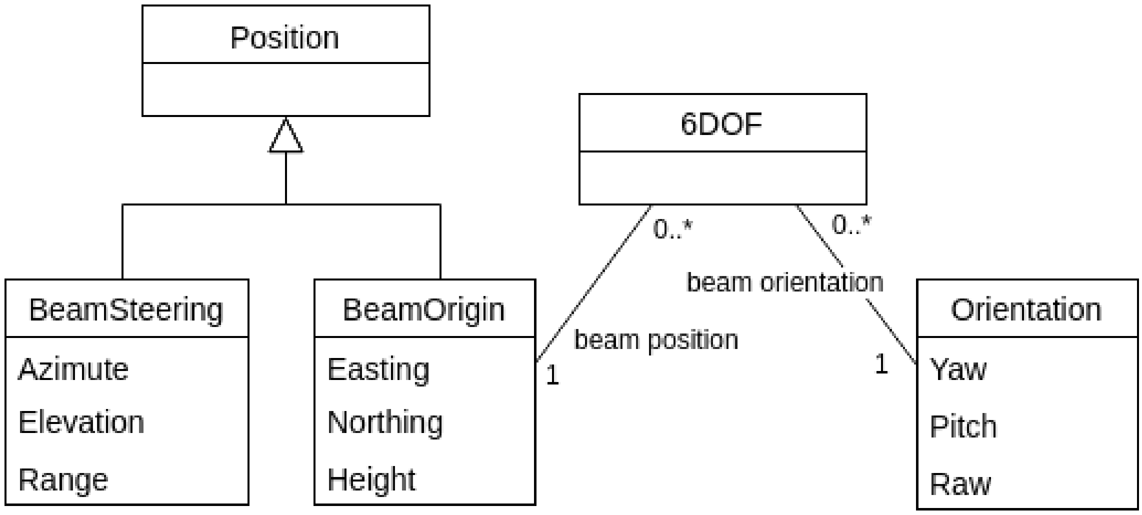


Figure 5.5: UML representation of positions and coordinate systems

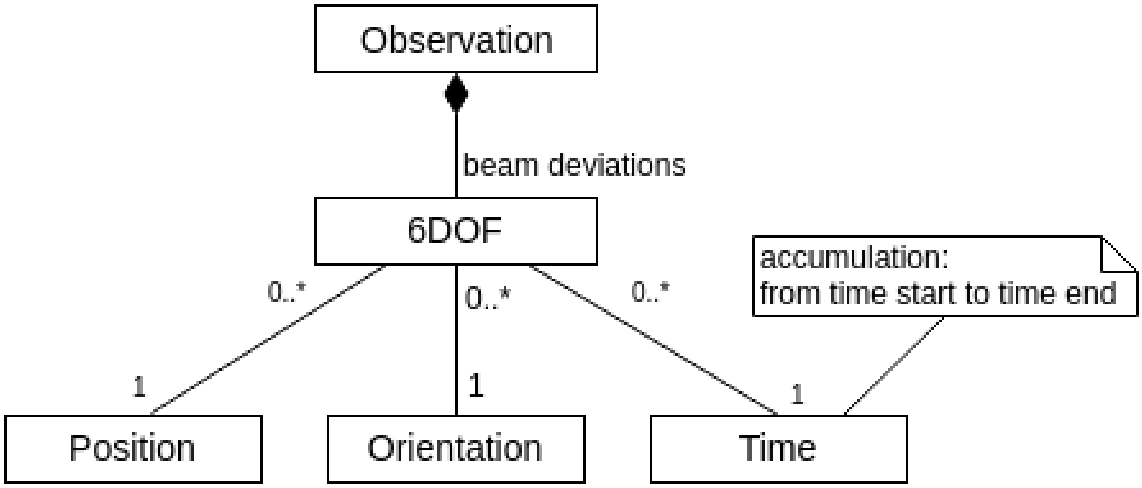


Figure 5.6: UML representation of the accumulation time for one measurement point

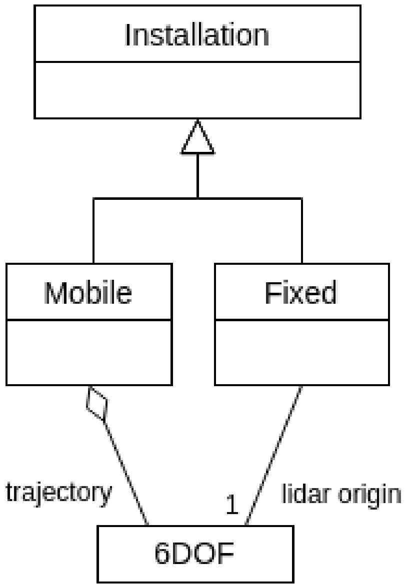


Figure 5.7: UML representation of the installation type

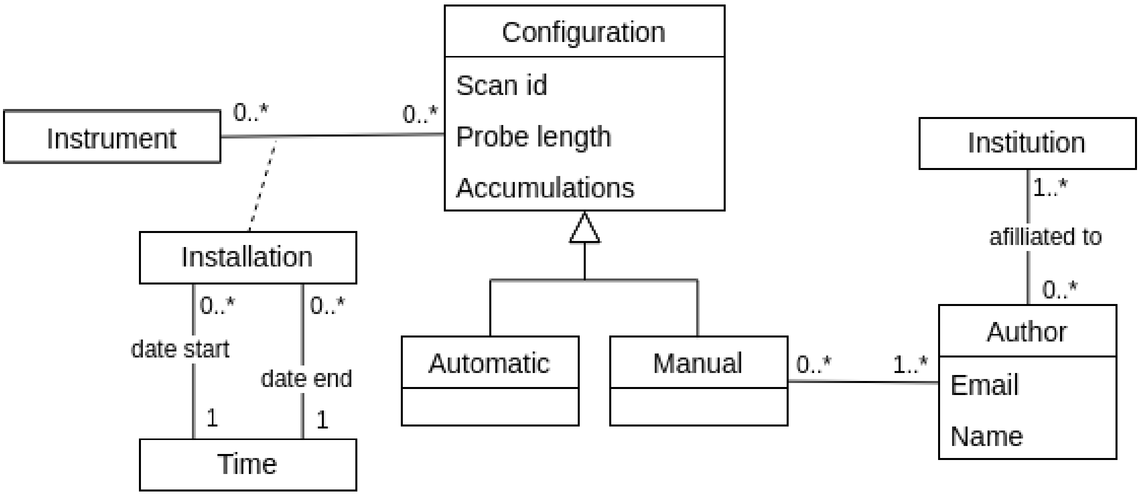


Figure 5.8: UML representation of the instrument configuration

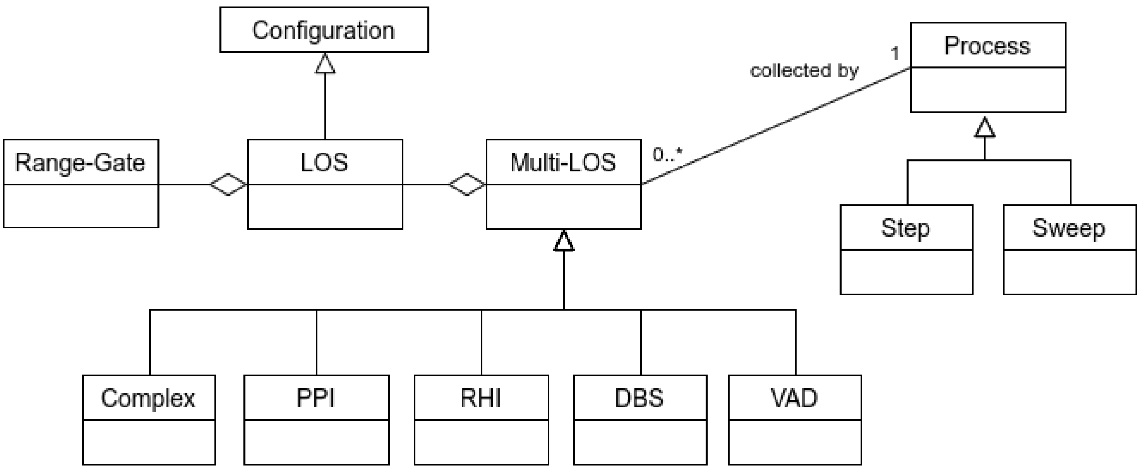


Figure 5.9: UML representation for different configurations

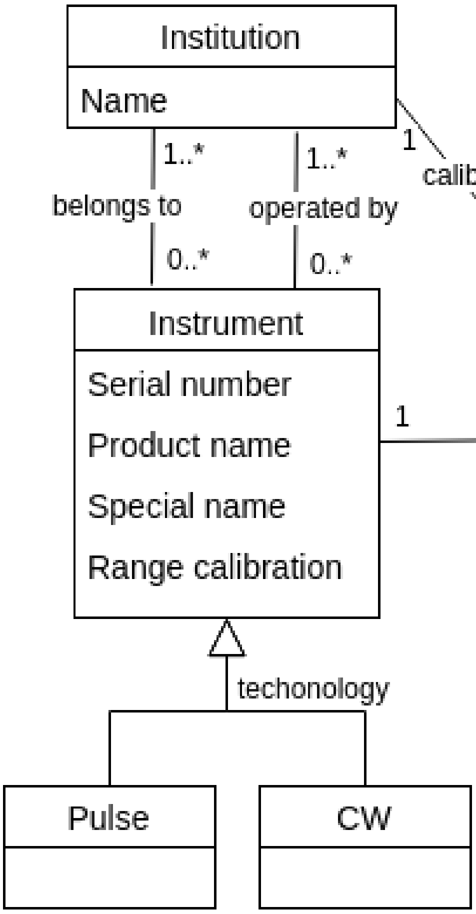


Figure 5.10: UML representation of Instruments and calibrations

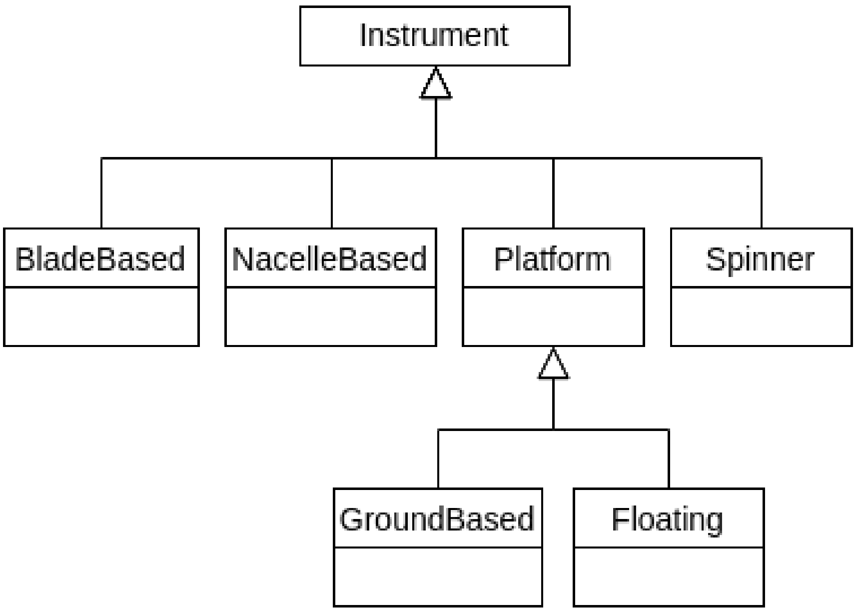


Figure 5.11: UML representation instrument installation type

Chapter 6

Making lidar data interoperable

In the previous section we have presented the conceptual model of L2 data product. Based on this description in this section we will select suitable file format to map our conceptual model. This section will in details describe previously established metadata and translate them to specific descriptors that will be stored in the selected format. Also, we will propose the approach how the measured features of interest should be organized in the files.

6.1 NetCDF overview

Up to this point, the conceptual model included the main concepts of the domain (i.e., wind lidars) related with the processes and the entities involved in the observations of features of interest. The stakeholders are now able to understand better the descriptors to be included in the file containing the L2 lidar data product.

The L2 output file, to be detailed in the following sections, will include the descriptors needed for researchers to better understand the process, the instruments used, the provenance, etc. as well as the data (values measured).

Figure 6.1 (based on the NetCDF4/CDM[1]) describes, in UML, the structure of the files used to encode L2 datasets and associated metadata.

[1] UCAR, NetCDF documentation — Enhanced Data Model: : https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_set_components.html

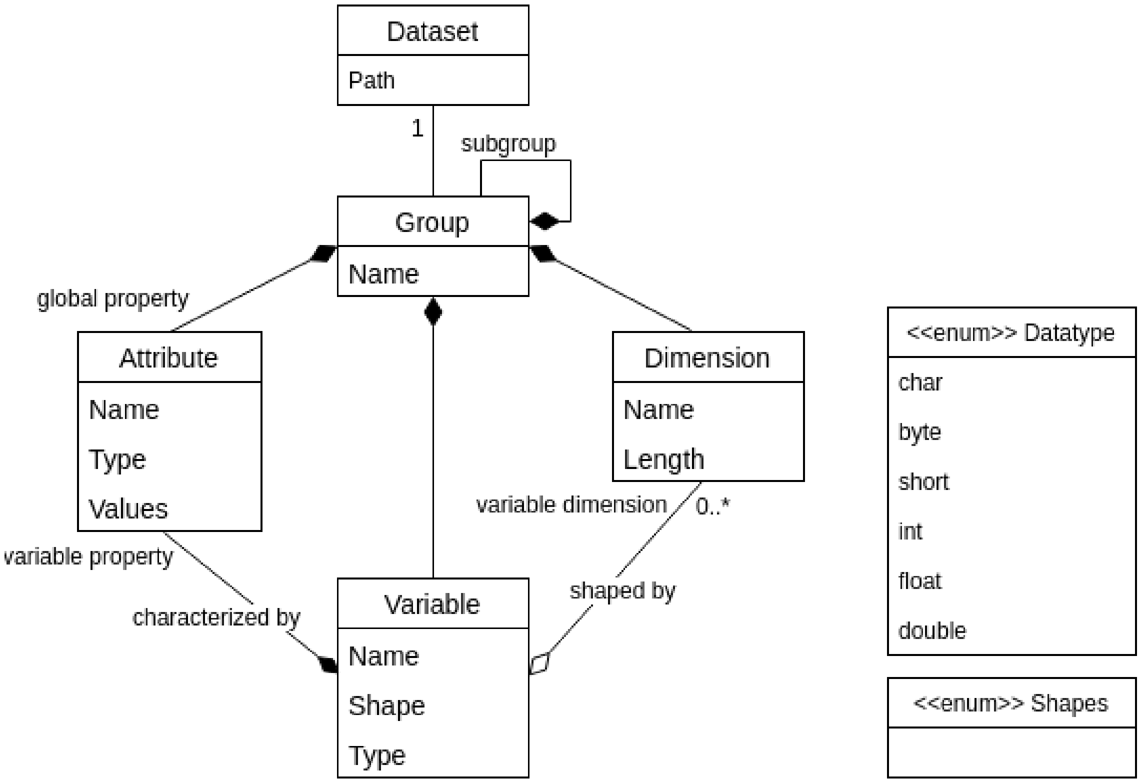


Figure 6.1: UML representation of the simple model for L2 datasets

A Dataset is stored in a file, with a known path, that contains one Group, which may contain subgroups (defined by scan-id).

Each group collects a set of observations (features of interest: radial velocity, CNR, etc.) stored in Variable. A variable holds the value of an observed feature of interest and has a name (radial velocity), a shape (time-range), and a type (float).

A Dimension may be used to represent a real physical dimension, for example, time, azimuth, elevation, or range. In case of a LOS configuration, the dimension range has a length equal to the maximum number of range gates that any of the recorded LOS measurements has.

The shape of a variable depends on the dimensions. It's specified (shaped by) accordingly to the dimensions of the variable: scalar if it has no dimension associated, vector for 1 dimension, matrix or grid for 2 dimension, etc.

An Attribute holds metadata (global descriptors or variable descriptors) and has a name, a type of Datatype (usually a string) and its value.

An attribute may describe groups or variables. It can hold global properties, if it is about the group (installation = LOS), or it holds variable properties for variables (units = m/s , longname = cdf-radial name).

6.2 Variables

Variables will be described by means of the following template:

Variable name:	*The variable name shall be written here. The name should provide an unambiguous notation of the variable which should directly imply the purpose of the variable. Names are case sensitive.*
Variable category*:	*The category to which the variable belongs will be stated in this field.*
Convention:	*Indicates in which convention the variable is used. If the variable is used in several conventions all conventions will be stated.*
Description*:	*A short description of the variable.*
Long name*:	*The variable long name shall be written here.*
Dimensions*:	*Indicates the dimensions of the variable.*
Type*:	*Indicates the type of the variable values.*
Units*:	*Indicates the units of the variable values.*
Mandatory:	*This field indicates whether the variable is mandatory in the e-WindLidar convention or not.*
Example:	*A practical example of the usage of the variable. In case when the variable is mandatory, the example should provide rations for requiring data creators to record this variable.*

The elements of this template marked with asterisk "*", such as *Long name*, represent attributes of variables which are recommended to be stored in the NetCDF file since this will improve the reuse of data.

6.2.1 Coordinate variables

Table 6.1: Variable *time*

Variable name:	*time*
Variable category:	*Coordinate variables*
Convention:	*CF/Radial , e-WindLidar*
Description:	*Time of the beginning of the averaging period in seconds since the experiment start date. Format:* $yyyy - mm - ddThh : mm : ssZ$
Long name:	*start_time_of_averaging_period*
Dimensions:	*time*
Type:	*double*
Units:	*seconds*
Mandatory:	*Yes*
Example:	

Table 6.2: Variable *range*

Variable name:	*time*
Variable category:	*Coordinate variables*
Convention:	*CF/Radial , e-WindLidar*
Description:	*Distance of the center of the range gate form the position of the lidar*
Long name:	*range_gate_distance_from_lidar*
Dimensions:	*range, time*
Type:	*double*
Units:	*meter*
Mandatory:	*Yes*
Example:	

6.2.2 Beam steering and location variables

Table 6.3: Variable *azimuth_angle*

Variable name:	*azimuth_angle*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*azimuth angle of the lidar beam at the start of the measurement, clockwise from North*
Long name:	*azimuth_angle_of_lidar_beam*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*Yes*
Example:	

Table 6.4: Variable *azimuth_angle_sweep*

Variable name:	*azimuth_angle_sweep*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*azimuth sector that is swept during the accumulation with positive angles turning right*
Long name:	*azimuth_sector_swept_during_accumulation*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*No*
Example:	

Table 6.5: Variable *elevation_angle*

Variable name:	*elevation_angle*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*Elevation angle of the lidar beam at the start of the measurement*
Long name:	*elevation_angle_of_lidar_beam*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*Yes*
Example:	

Table 6.6: Variable *elevation_angle_sweep*

Variable name:	*elevation_angle_sweep*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*Elevation sector that is swept during the accumulation with positive angles turning from horizon to the sky*
Long name:	*elevation_sector_swept_during_accumulation*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*No*
Example:	

Table 6.7: Variable *yaw*

Variable name:	*yaw*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*yaw angle of the lidar device*
Long name:	*idar_yaw_angle*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*Yes*
Example:	

Table 6.8: Variable *pitch*

Variable name:	*pitch*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*pitch angle of the lidar device*
Long name:	*idar_pitch_angle*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*Yes*
Example:	

Table 6.9: Variable *roll*

Variable name:	*roll*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*roll angle of the lidar device*
Long name:	*idar_roll_angle*
Dimensions:	*none or time*
Type:	*double*
Units:	*degrees*
Mandatory:	*Yes*
Example:	

Table 6.10: Variable *position_x*

Variable name:	*position_x*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*x coordinate of the lidar, unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*x_position_of_lidar*
Dimensions:	*none or time*
Type:	*double*
Units:	*Meters or degrees*
Mandatory:	*Yes*
Example:	

Table 6.11: Variable *position_y*

Variable name:	*position_y*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*y coordinate of the lidar, unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*y_position_of_lidar*
Dimensions:	*none or time*
Type:	*double*
Units:	*Meters or degrees*
Mandatory:	*Yes*
Example:	

Table 6.12: Variable *position_z*

Variable name:	*position_z*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*z coordinate of the lidar, unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*z_position_of_lidar*
Dimensions:	*none or time*
Type:	*double*
Units:	*Meters or degrees*
Mandatory:	*Yes*
Example:	

Table 6.13: Variable *moving_speed_x*

Variable name:	*moving_speed_x*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Speed of the lidar along x coordinate. Unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*x_speed_of_lidar*
Dimensions:	*time, time and range (CW lidars)*
Type:	*double*
Units:	$m.s^{-1}$ *or* $degrees.s^{-1}$
Mandatory:	*No*
Example:	

Table 6.14: Variable *moving_speed_y*

Variable name:	*moving_speed_y*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Speed of the lidar along y coordinate. Unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*y_speed_of_lidar*
Dimensions:	*time, time and range (CW lidars)*
Type:	*double*
Units:	$m.s^{-1}$ *or* $degrees.s^{-1}$
Mandatory:	*No*
Example:	

Table 6.15: Variable *moving_speed_z*

Variable name:	*moving_speed_z*
Variable category:	*Beam steering and location variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Speed of the lidar along z coordinate. Unit depends on a selected coordinate system. See the global attribute coordinate_system.*
Long name:	*z_speed_of_lidar*
Dimensions:	*time, time and range (CW lidars)*
Type:	*double*
Units:	$m.s^{-1}$ *or* $degrees.s^{-1}$
Mandatory:	*No*
Example:	

Table 6.16: Variable *acceleration_x*

Variable name:	*acceleration_x*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*Acceleration of the lidar along x coordinate. Unit depends on a selected coordinate system. See the global attribute coordinate_system.*
Long name:	*x_acceleration_of_lidar*
Dimensions:	*time, time and range (CW lidars)*
Type:	*double*
Units:	$m.s^{-2}$ *or* $degrees.s^{-2}$
Mandatory:	*No*
Example:	

Table 6.17: Variable *acceleration_y*

Variable name:	*acceleration_y*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*Acceleration of the lidar along y coordinate. Unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*y_acceleration_of_lidar*
Dimensions:	*time, time and range (CW lidars)*
Type:	*double*
Units:	$m.s^{-2}$ *or* $degrees.s^{-2}$
Mandatory:	*No*
Example:	

Table 6.18: Variable *acceleration_z*

Variable name:	*acceleration_z*
Variable category:	*Beam steering and location variables*
Convention:	*e-WindLidar*
Description:	*Acceleration of the lidar along z coordinate. Unit depends on a selected coordinate system. See the global attribute coordinate_system*
Long name:	*z_acceleration_of_lidar*
Dimensions:	*time, time and range (CW lidars)*
Type:	*double*
Units:	$m.s^{-2}$ *or* $degrees.s^{-2}$
Mandatory:	*No*
Example:	

6.2.3 Data variables

Table 6.19: Variable *scan_type*

Variable name:	*scan_type*
Variable category:	*Data variables*
Convention:	*e-WindLidar*
Description:	*This variable indicates type of the scanning configuration. We use the following encoding for different type of scanning types: 0 - other, 1 - starring, 2 - DBS, 3 - VAD, 4 - PPI, 5 - RHI. In case we indetify new scanning types we will extend this list with new ids.*
Long name:	*scan_type_of_the_measurement*
Dimensions:	*None or time*
Type:	*integer*
Units:	*none*
Mandatory:	*Yes*
Example:	

Table 6.20: Variable *scan_id*

Variable name:	*scan_id*
Variable category:	*Data variables*
Convention:	*e-WindLidar*
Description:	*This variable indicates id of the scanning configuration. For example if the file contains several different PPI scans, then using this variable end users will be able to distinguish data that belongs to specific PPI configurations.*
Long name:	*scan_id_of_the_measurement*
Dimensions:	*None or time*
Type:	*integer*
Units:	*none*
Mandatory:	*No*
Example:	

Table 6.21: Variable *accumulation_time*

Variable name:	*accumulation_time*
Variable category:	*Data variables*
Convention:	*e-WindLidar*
Description:	*The time interval for the accumulation of Doppler spectra.*
Long name:	*time_for_spectral_accumulation*
Dimensions:	*None or time or time and range*
Type:	*double*
Units:	*seconds*
Mandatory:	*yes*
Example:	

Table 6.22: Variable *n_spectra*

Variable name:	*n_spectra*
Variable category:	*Data variables*
Convention:	*e-WindLidar*
Description:	*This variable indicates number of spectra used to produce the averaged spectra which is then processed to estimate radial velocity.*
Long name:	*number_of_spectra*
Dimensions:	*None or time or time and range*
Type:	*integer*
Units:	*seconds*
Mandatory:	*no*
Example:	

6.2.4 Measurement variables

Table 6.23: Variable *VEL*

Variable name:	*VEL*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Radial velocity.*
Long name:	*radial_velocity*
Dimensions:	*Time and range*
Type:	*double*
Units:	$m.s^{-1}$
Mandatory:	*yes*
Example:	

Table 6.24: Variable *CNR*

Variable name:	*CNR*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Carrier to noise ratio*
Long name:	*carrier_to_noise_ratio*
Dimensions:	*Time and range*
Type:	*double*
Units:	*dB*
Mandatory:	*yes*
Example:	

Table 6.25: Variable *WIDTH*

Variable name:	*WIDTH*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Doppler spectrum width*
Long name:	*Doppler_spectrum_width*
Dimensions:	*Time and range*
Type:	*double*
Units:	$m.s^{-1}$
Mandatory:	*no*
Example:	

Table 6.26: Variable *T*

Variable name:	*T*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*External temperature measured by met station sensors*
Long name:	*met_station_temperature*
Dimensions:	*Time*
Type:	*double*
Units:	*Degrees C*
Mandatory:	*no*
Example:	

Table 6.27: Variable *T_internal*

Variable name:	*T_internal*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Temperature measured inside of lidar*
Long name:	*internal_lidar_temperature*
Dimensions:	*Time*
Type:	*double*
Units:	*Degrees C*
Mandatory:	*no*
Example:	

Table 6.28: Variable *P*

Variable name:	*P*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Barometric pressure measured by met station sensors*
Long name:	*met_station_pressure*
Dimensions:	*Time*
Type:	*double*
Units:	*Pa*
Mandatory:	*no*
Example:	

Table 6.29: Variable *P_internal*

Variable name:	*P_internal*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Barometric pressure inside of lidar*
Long name:	*internal_pressure*
Dimensions:	*Time*
Type:	*double*
Units:	*Pa*
Mandatory:	*no*
Example:	

Table 6.30: Variable *RH*

Variable name:	*RH*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Relative humidity measured by met station sensors*
Long name:	*met_station_humidity*
Dimensions:	*Time*
Type:	*double*
Units:	*percent*
Mandatory:	*no*
Example:	

Table 6.31: Variable *RH_internal*

Variable name:	*RH_internal*
Variable category:	*Measurement variables*
Convention:	*CF/Radial, e-WindLidar*
Description:	*Relative humidity measured inside of the lidar*
Long name:	*lidar_humidity*
Dimensions:	*Time*
Type:	*double*
Units:	*percent*
Mandatory:	*no*
Example:	

6.2.5 Storing matrices of variables

Variables are stored in matrix form in the NetCDF file. Depending on the actual measurement setup these matrices can either have no dimension, only time dimension or time and range dimension.

Time and range are the only dimensions used in the e-WindLidar NetCDF format and are also called "coordinate variables" in accordance to the CF/radial convention. Measurement variables have both dimensions which results in a 2D matrix for each variable. In case of varying range gates with time, the range variable has the length of the maximum number of range gates over all time steps in the data file (or the data set from that measurement campaign in case of split data files). Empty range gates are filled up with 'NaN' values.

Some variables can either have no dimension or time dimension. E.g. in case of an RHI measurement scenario the azimuth angle does not change and can thus be stored without dimension. For the same case, the elevation does change with time and is therefore assigned time as dimension, resulting in a 1D matrix. In case of more complex measurement scenarios both, elevation and azimuth can have time dimension.

Many variables are constant for the whole data set and do not have any dimension at all. Common examples are position_x, position_y, position_z or pitch, yaw, roll, which usually do not change for a ground based lidar measurement.

Nevertheless, most of the variables could possibly change with time depending on the complexity of the measurement scenario (e.g. changing accumulation time) or the installation_type (e.g. floating lidars). The e-WindLidar convention allows for this making it applicable to most lidar applications.

Figure 6.2 is exemplary showing a typical representation of common variables in the NetCDF format for the e-WindLidar convention. Dimensions and coordinate variables are marked in blue. In this example elevation and azimuth angles are shown as 1D matrices which might be the case for complex trajectories. On the other hand, the location and data variables are shown as constant variables.

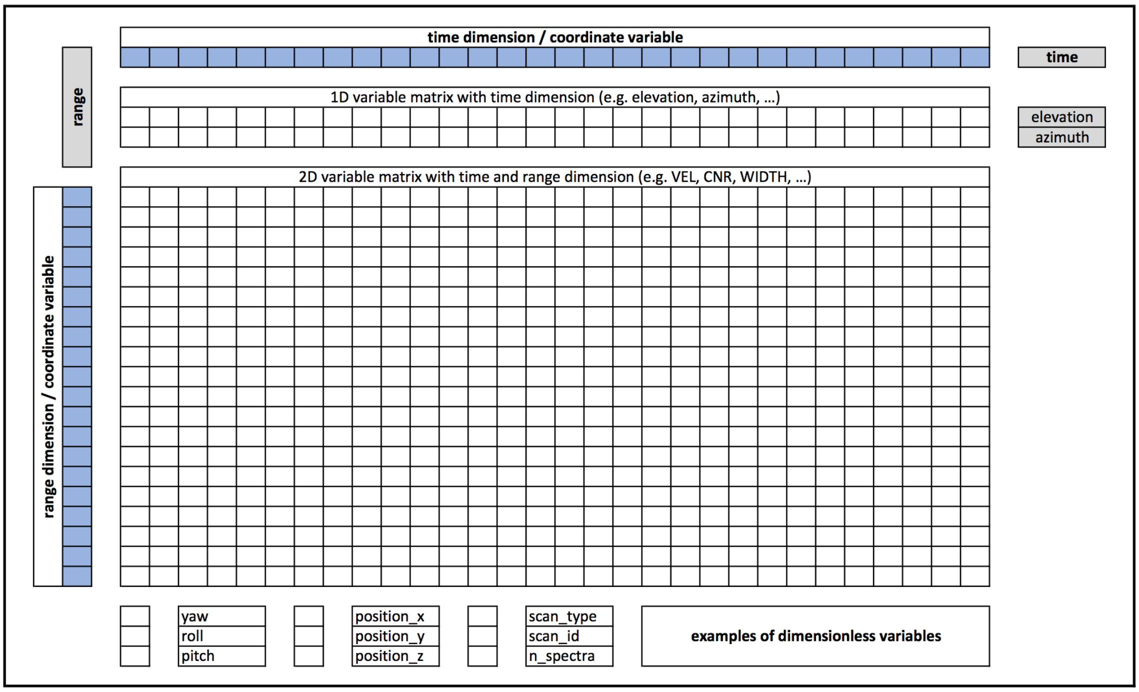


Figure 6.2: Graphical representation how variables are stored in the NetCDF file

6.3 Attributes

Attributes will be described by means of the following template:

Attribute name:	*The attribute name shall be written here. The name should provide an unambiguous notation of the attribute which should directly imply the purpose of the attribute. Names are case sensitive.*
Attribute category:	*The category to which the attribute belongs will be stated in this field.*
Convention:	*Indicates in which convention the attribute is used. If the attribute is used in several conventions all conventions will be stated.*
Description:	*A short description of the attribute.*
Type:	*Indicates the type of the attrubute value(s).*
Values:	*Indicates either a range of values or restricted set of values that the attribute can take.*
Format:	*Indicates whether the values are recorded in simple or complex format.*
Structure:	*In case of complex formatting the formatting structure is provided. Otherwise in this field value None will be written.*
Mandatory:	*This field indicates whether the attribute is mandatory in the e-WindLidar convention or not.*
Example:	*A practical example of the usage of the attribute. In case when the attribute is mandatory, the example should provide rations for requiring data creators to record this attribute.*

6.3.1 General attributes

Table 6.32: Attribute *conventions*

Attribute name:	*conventions*
Attribute category:	*General attribute*
Convention:	*CF/Radial, e-WindLidar*
Description:	*This attribute indicates the convention that is used to structure and record lidar data in NetCDF format.*
Type:	*string*
Values:	*CF, CF/Radial, e-WindLidar*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*When conventions is equal to e-WindLidar, NetCDF files will have a characteristic structure with a set of attributes and variables which are defined by this convention. The value of this attribute indicates to the data user what convention was used to structure and record lidar data. Moreover, this attribute can in future represent an input parameter for algorithms which are processing NetCDF files.*

Table 6.33: Attribute *version*

Attribute name:	*version*
Attribute category:	*General attribute*
Convention:	*CF/Radial, e-WindLidar*
Description:	*This attribute is used to indicate the version of the convention that is used to structure and record lidar data in NetCDF format.*
Type:	*string*
Values:	*Any character or combination of characters*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*Considering that the attribute conventions is set to CF/Radial (i.e., conventions = CF/Radial) the attribute version can take value equal to 1 or 2 (e.g., version = 1). Currently, if the attribute conventions is set to e-WindLidar, the version attribute can take only value 1.0 . The main reason for having this attribute mandatory is to differentiate among versions of the conventions since some algorithms and services are specifically developed to work with data recorded using a certain convention and certain version of that convention.*

Table 6.34: Attribute *title*

Attribute name:	*title*
Attribute category:	*General attribute*
Convention:	*CF/Radial, e-WindLidar*
Description:	*This attribute indicates the name by which the dataset is formally known. Typically the attribute will take the name of the experiment in which the given dataset was recorded.*
Type:	*string*
Values:	*Any character or combination of characters*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*In case of the dataset recorded during the Kassel campaign, which take place in 2014, the attribute title will have a value equal to Kassel-2015 (i.e., title = Kassel-2014). Giving the title to a dataset represents a way to relate the dataset to a measurement campaign and to differentiate among other datasets created in different measurement campaigns*

Table 6.35: Attribute *creator*

Attribute name:	*creator*
Attribute category:	*General attribute*
Convention:	*e-WindLidar*
Description:	*This attribute points to the creator(s) of the dataset.*
Type:	*string*
Values:	*Names of data creators and their institutions.*
Format:	*Complex*
Structure:	*CreatorName_1 - CreatorInstitute_1,..., CreatorName_N - CreatorInstitute_N*
Mandatory:	*Yes*
Example:	*In case of the Perdigao-2015 experiment the attribute creator will have following values:* *creator = Nikola Vasiljevic - Technical University of Denmark, Nikolas Angelou - Technical University of Denmark, Guillaume Lea - Technical University of Denmark, Robert Menke - Technical University of Denmark.* *This attribute is mandatory since it provides the visibility of data creators.*

Table 6.36: Attribute *references*

Attribute name:	*references*
Attribute category:	*General attribute*
Convention:	*e-WindLidar, CF/Radial*
Description:	*Unambiguous reference(s) to the resource which describes the dataset. This can be report, journal publication or similar. Preferably this attribute should contain a persistent identifier to the resource (e.g., DOI).*
Type:	*String*
Values:	*Resource title, resources authors names, web links to the resource, persistent identifier to the resource, etc.*
Format:	*Simple*

Structure:	`Reference_1`
	`<empty row>`
	`Refernce_2`
	`<empty row>`
	`Reference_3`
	`<empty row>`
	`...`
	`<empty row>`
	`Reference_N`

Mandatory:	*Yes (if there is a resource that describes the dataset)*
Example:	*Providing a reference to the resource that describes the dataset helps data users to better interpret data, since the resource will provide additional explanation of the dataset which are not typically recorded in the dataset (e.g., what type of flow conditions were intended to be recorded). In case of several references, two consecutive references should be separated with an empty row.* *In case of Perdigao-2015 dataset the attribute references, containing two information sources, will have the following value:* *references = Vasiljevic, N., L. M. Palma, J. M., Angelou, N., Carlos Matos, J., Menke, R., Lea, G., Mann, J., Courtney, M., Frolen Ribeiro, L., and M. G. C. Gomes, V. M.: Perdigao 2015: methodology for atmospheric multi-Doppler lidar experiments, Atmos. Meas. Tech., 10, 3463-3483* https://doi.org/10.5194/amt-10-3463-2017, *2017* http://perdigao-2015.tumblr.com

Table 6.37: Attribute *site*

Attribute name:	*site*
Variable category:	*General attribute*
Convention:	*e-WindLidar, CF/Radial*
Description:	*The name of the site where the measurement campaign took place during which the given dataset was recorded.*
Type:	*String*
Values:	*Any character or combination of characters*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*In case of Perdigão-2015 dataset the attribute site will have the following value:* *site = Serra do Perdigão*

Table 6.38: Attribute *general_comment*

Attribute name:	*general_comment*
Variable category:	*General attribute*
Convention:	*e-WindLidar*
Description:	*Additional general comments that are not related to the instrument used in the campaign, the measurement configuration or to the data and that do not fit into any of the defined attributes should be provided via this attribute.*
Type:	*String*
Values:	*Comments should be written in English.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*This attribute can contain a range of comments related to for example the site where the measurement campaign took place.*

Table 6.39: Attribute *general_comment_xml*

Attribute name:	*general_comment_xml*
Variable category:	*General attribute*
Convention:	*e-WindLidar*
Description:	*Additional general comments that are not related to the instrument used in the campaign, the measurement configuration or to the data and that do not fit into any of the defined attributes should be provided via this attribute.*
Type:	*String*
Values:	*Comments should be written in English.*
Format:	*Complex*
Structure:	<LIST_OF_COMMENTS> <COMMENT> <ID>1</ID> <DATE_TIME>YYYY-MM-DDThh:mmTZD</DATE_TIME> <AUTHOR>First_Name Last_Name</AUTHOR> <MESSAGE>Comment</MESSAGE> </COMMENT> ... <COMMENT> <ID>N</ID> <DATE_TIME>YYYY-MM-DDThh:mmTZD</DATE_TIME> <AUTHOR>First_Name Last_Name</AUTHOR> <MESSAGE>Comment</MESSAGE> </COMMENT> </LIST_OF_COMMENTS>
Mandatory:	*No*
Example:	<LIST_OF_COMMENTS> <COMMENT> <ID>1</ID> <DATE_TIME>2017-06-23T08:34T+01:00</DATE_TIME> <AUTHOR>Lukas Pauscher</AUTHOR> <MESSAGE>Everything works just fine.</MESSAGE> </COMMENT> <COMMENT> <ID>2</ID> <DATE_TIME>2017-06-23T10:34T+01:00</DATE_TIME> <AUTHOR>Nikola Vasiljevic</AUTHOR> <MESSAGE>Scanner head stopped.</MESSAGE> </COMMENT> </LIST_OF_COMMENTS>

6.3.2 Instrument attributes

Table 6.40: Attribute *lidar_technology*

Attribute name:	*lidar_technology*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates the lidar technology.*
Type:	*String*
Values:	*Pulsed, Continuous Wave*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*The indication of the lidar technology is essential since there are differences between CW and pulsed lidars, such as probe length, measurement rate, data filtering, etc.. If, for example, ZephIR was used to produce data, this attribute will be set to Continuous Wave*

Table 6.41: Attribute *lidar_scanning_type*

Attribute name:	*lidar_technology*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates the scanning type of lidar.*
Type:	*String*
Values:	*pointing, horizontal profiling, vertical profiling, scanning*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*The scanning type indicates capabilities a lidar has in term of scanning geometries. If, for example, Windcube V2 was used to produce data, this attribute will have value set to "vertical profiling"*

Table 6.42: Attribute *lidar_range_classification*

Attribute name:	*lidar_technology*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates a lidar's capability in terms of range.*
Type:	*String*
Values:	*ultra-short, short, medium, long and ultra-long*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*If, for example, Windcube V2 was used to produce data, this attribute will have the value set to "short".*

Table 6.43: Attribute *lidar_installation_type*

Attribute name:	*lidar_installation_type*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute further refines description of the attribute lidar_is_mobile by indicating the installation type of a lidar.*
Type:	*String*
Values:	*ground-based, floating, nacelle-based, blade-based, spinner*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*If, for example, Windcube V2 was used to produce data, this attribute will have the value set to "ground-based".*

Table 6.44: Attribute *product_name*

Attribute name:	*product_name*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates the lidar product name.*
Type:	*String*
Values:	*Product names of the lidar given by their producers.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*If, for example, Leosphere horizontal profiler was used to produce data, this attribute might take value "Windcube V2".*

Table 6.45: Attribute *lidar_is_mobile*

Attribute name:	*product_name*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attributes indicates whether the lidar position, orientation and leveling were fixed during the measurement campaign or a lidar was not stationary.*
Type:	*Boolean*
Values:	*yes, no*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*In case of the turbine mounted lidar or a lidar on a buoy this attribute will have value set to "yes".*

Table 6.46: Attribute *lidar_orientation*

Attribute name:	*product_name*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates the orientation of a lidar installed at the nacelle of a wind turbine (either upwind or downwind).*
Type:	*String*
Values:	*Upwind or Downwind*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*Values are self-explanatory*

Table 6.47: Attribute *flow_direction_encoding*

Attribute name:	*product_name*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates how the sign radial wind speed should be interpreted*
Type:	*Integer*
Values:	*0 or 1*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*If flow_direction_encoding is set to 0, this means if the wind is approaching the lidar the measured radial velocity will be recorded with negative sign. Otherwise, if flow_direction_encoding is set to 1, the recorded radial velocity will be stored with plus sign.*

Table 6.48: Attribute *serial_number*

Attribute name:	*serial_number*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute holds the serial number of the lidar used to produce the dataset.*
Type:	*String*
Values:	*Any character and/or combination of characters*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No if specific_lidar_name is provided, otherwise Yes*
Example:	*In case if specific_lidar_name is not provided this attribute is the only source of information that can be used to identify a lidar used in the measurement campaign.*

Table 6.49: Attribute *specific_lidar_name*

Attribute name:	*specific_lidar_name*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute holds information about a specific name given to the lidar.*
Type:	*String*
Values:	*Any character and/or combination of characters*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No if serial_number is provided, otherwise Yes*
Example:	*DTU Wind Energy uses specific names for WindScanners (scanning lidars). In case of the long-range WindScanner the specific names correspond to well known winds from different countries. For example, one of the long-range WindScanners is named Koshava.*

Table 6.50: Attribute *lidar_owner*

Attribute name:	*lidar_owner*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates the owner of the lidar that recorded the dataset.*
Type:	*String*
Values:	*Any combination of characters which indicates the name of a party that owns the lidar.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*lidar_owner = UPORTO*

Table 6.51: Attribute *lidar_operator*

Attribute name:	*lidar_operator*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates the operator of the lidar that recorded the dataset.*
Type:	*String*
Values:	*Any combination of characters which indicates the name of a party that operates the lidar.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*lidar_operator = NTNU*

Table 6.52: Attribute *instrument_comment*

Attribute name:	*instrument_comment*
Variable category:	*Instrument attributes*
Convention:	*e-WindLidar*
Description:	*This attribute holds comments related to the lidar that can help in data analysis.*
Type:	*String*
Values:	*Comments should be written in English.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*This can be a simple text stating for example whether the lidar had a malfunction during the measurement campaign which can explain for an empty period in the data record.*

Table 6.53: Attribute *instrument_comment_xml*

Attribute name:	*instrument_comment_xml*
Variable category:	*General attribute*
Convention:	*e-WindLidar*
Description:	*This attribute holds comments related to the lidar that can help in data analysis*
Type:	*String*
Values:	*Comments should be written in English.*
Format:	*Complex*

Structure:	`<LIST_OF_COMMENTS>`
	`<COMMENT>`
	`<ID>1</ID>`
	`<DATE_TIME>YYYY-MM-DDThh:mmTZD</DATE_TIME>`
	`<AUTHOR>First_Name Last_Name</AUTHOR>`
	`<MESSAGE>Comment</MESSAGE>`
	`</COMMENT>`
	`...`
	`<COMMENT>`
	`<ID>N</ID>`
	`<DATE_TIME>YYYY-MM-DDThh:mmTZD</DATE_TIME>`
	`<AUTHOR>First_Name Last_Name</AUTHOR>`
	`<MESSAGE>Comment</MESSAGE>`
	`</COMMENT>`
	`</LIST_OF_COMMENTS>`

Mandatory:	*No*

Example:	`<LIST_OF_COMMENTS>`
	`<COMMENT>`
	`<ID>1</ID>`
	`<DATE_TIME>2015-05-26T12:00T+1</DATE_TIME>`
	`<AUTHOR>Nikola Vasiljevic</AUTHOR>`
	`<MESSAGE>Measurements acquired during 23,`
	`24 and 25 of May should not be used since`
	`Koshava was pointing in a wrong direction.`
	`</MESSAGE>`
	`</COMMENT>`
	`</LIST_OF_COMMENTS>`

6.3.3 Measurement configuration attributes

Table 6.54: Attribute *n_gates_vary*

Attribute name:	*n_gates_vary*
Variable category:	*Measurement configuration attributes*
Convention:	*CF Radial, e-WindLidar*
Description:	*This attribute indicates whether the number of range gates is fixed for all LOS measurements.*
Type:	*Boolean*
Values:	*yes, no*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*For commercial lidars, such as Windcube V2, number of range gates is fixed. Therefore, for these lidars the attribute n_gates_vary will be set to "no".*

Table 6.55: Attribute *spatial_averaging_info*

Attribute name:	*spatial_averaging_info*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attribute provides information on the spatial averaging within the range gate.*
Type:	*String*
Values:	*Numbers and/or formula*
Format:	*Simple*
Structure:	*Scan_ID_1: Probe_Length_1_Info, ..., Scan_ID_N: Probe_Length_N_Info*
Mandatory:	*No*
Example:	*In case of pulsed lidars, where the NetCDF file contains only data from a single measurement configuration, this attribute will have a fixed value, which will represent a full-width half maximum of a total probe volume (i.e., effective probe length). If a CW lidar is used to record data this attribute should hold a formula that provides means to calculate the effective probe length with respect to the focusing distance.*

Table 6.56: Attribute *beam_sweeping*

Attribute name:	*beam_sweeping*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attributes indicate whether the beam is moving with the respect to the beam origin during the spectra accumulation, which directly indicates whether besides the longitudinal spatial averaging (effective probe length) also lateral spatial averaging occurs.*
Type:	*Boolean*
Values:	*yes, no*
Format:	*Simple*
Structure:	*None*
Mandatory:	*yes*
Example:	*During the spectra accumulation of non-scanning Windcubes (pulsed lidars) the beam position with respect to the beam origin is fixed, which means that the attribute beam_sweeping will have value equal to no. On the other hand, ZephIR lidars (CW lidars) continuously move the laser beam during the spectra accumulation.*

Table 6.57: Attribute *measurement_scenario*

Attribute name:	*measurement_scenario*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attribute contains description of the measurement scenario(s) used to configure the lidar to perform measurements that are recorded in the current NetCDF file.*
Type:	*String*
Values:	*Free text*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes, unless the attribute measurement_scenario_xml is provided*
Example:	*The data creators should provide sufficiently comprehensive information about the measurement scenario(s). If one NetCDF file contains data recorded by application of more than one measurement scenario it is essential to link the description of each measurement scenario to the ScanID and ScanType variables.*

Table 6.58: Attribute *measurement_scenario_xml*

Attribute name:	*measurement_scenario*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attribute contains description of the measurement scenario(s) used to configure the lidar to perform measurements that are recorded in the current NetCDF file.*
Type:	*String*
Values:	
Format:	*Complex*
Structure:	<LIST_OF_SCENARIOS> <SCENARIO scan_id="1" scan_type="1-5" author="name"> <los los_id="1" FFT_size="" pulse_length="" azimuth_start="" elevation_start="" azimuth_stop="" elevation_stop="" accumulation_time="" transition_time="" range_gates="rg1, .. rgN"></los> ... <los los_id="N" FFT_size="" pulse_length="" azimuth_start="" elevation_start="" azimuth_stop="" elevation_stop="" accumulation_time="" transition_time="" range_gates="rg1, .. rgN"></los> </SCENARIO> ... <SCENARIO scan_id="N" scan_type="1-5" author="name"> <los los_id="1" FFT_size="" pulse_length="" azimuth_start="" elevation_start="" azimuth_stop="" elevation_stop="" accumulation_time="" transition_time="" range_gates="rg1, .. rgN"></los> ... <los los_id="N" FFT_size="" pulse_length="" azimuth_start="" elevation_start="" azimuth_stop="" elevation_stop="" accumulation_time="" transition_time="" range_gates="rg1, .. rgN"></los> </SCENARIO> </LIST_OF_SCENARIOS>
Mandatory:	*Yes, unless the attribute measurement_scenario is provided*
Example:	*Using this attribute the data creators can provide a detailed information on the measurement scenario(s).*

Table 6.59: Attribute *n_lidars*

Attribute name:	*measurement_scenario*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attribute indicates if the dataset is created while a lidar was operated in single- or multi- Doppler setup.*
Type:	*Integer*
Values:	*Positive integer values starting from 1 onwards.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*In case of Windcube V2 the value of this attribute will be set to 1, whereas in a typical operation of WindScanners this attribute will have value equal to 2 (dual-Doppler setup) or 3 (triple-Doppler setup).*

Table 6.60: Attribute *linked_lidars*

Attribute name:	*measurement_scenario*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*In case of multi-lidar setup, this attribute indicates to what other datasets/lidars the current dataset is connected.*
Type:	*String*
Values:	*Any notation for lidars which would allow straightforward identification of related datasets to the current one.*
Format:	*Simple*
Structure:	*lidar_name_1, lidar_name_2, ..., lidar_name_n*
Mandatory:	*Yes if n_lidars has value larger than one*
Example:	*In case of Perdigao-2015 campaign if the current dataset is recorded by the WindScanner named Koshava, the attribute linked_lidars would have value set to Sterenn, which represent specific lidar name of another WindScanner unit that was used in combination with Koshava to derive dual-Doppler measurements.*

Table 6.61: Attribute *configuration_comment*

Attribute name:	*configuration_comment*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attribute holds comments related to the lidar configuration that can help in data analysis.*
Type:	*String*
Values:	*Comments should be written in English.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*No*
Example:	*configuration_comment = The first measurement scenario is intended for single-Doppler retrievals. The second scenario was setup such that there in combination with the second lidar named Sterenn we were achieving true dual-Doppler measurements..*

Table 6.62: Attribute *configuration_comment_xml*

Attribute name:	*configuration_comment_xml*
Variable category:	*Measurement configuration attributes*
Convention:	*e-WindLidar*
Description:	*This attribute holds comments related to the lidar configuration that can help in data analysis.*
Type:	*String*
Values:	*Comments should be written in English.*
Format:	*Simple*
Structure:	<LIST_OF_COMMENTS> <COMMENT> <ID>1</ID> <DATE_TIME>YYYY-MM-DDThh:mmTZD</DATE_TIME> <AUTHOR>First_Name Last_Name</AUTHOR> <MESSAGE>Comment</MESSAGE> </COMMENT> ... <COMMENT> <ID>N</ID> <DATE_TIME>YYYY-MM-DDThh:mmTZD</DATE_TIME> <AUTHOR>First_Name Last_Name</AUTHOR> <MESSAGE>Comment</MESSAGE> </COMMENT> </LIST_OF_COMMENTS>
Mandatory:	*No*
Example:	<LIST_OF_COMMENTS> <COMMENT> <ID>1</ID> <DATE_TIME>2017-06-23T10:34T+01</DATE_TIME> <AUTHOR>Nikola Vasiljevic</AUTHOR> <MESSAGE>The first measurement scenario is intended for single-Doppler retrievals </MESSAGE> </COMMENT> </LIST_OF_COMMENTS>

6.3.4 Data description attributes

Table 6.63: Attribute *data_processing_history*

Attribute name:	*data_processing_history*
Variable category:	*Data description attributes*
Convention:	*e-WindLidar*
Description:	*This attribute states what processing data undergone from the lowest level to the current one.*
Type:	*String*
Values:	*Text should be written in English.*
Format:	*Simple*
Structure:	*None*
Mandatory:	*Yes*
Example:	*configuration_comment = The first measurement scenario is intended for single-Doppler retrievals. The second scenario was setup such that there in combination with the second lidar named Sterenn we were achieving true dual-Doppler measurements..*

Table 6.64: Attribute *data_processing_history_xml*

Attribute name:	*data_processing_history_xml*
Variable category:	*Data description attributes*
Convention:	*e-WindLidar*
Description:	*This attribute states what processing data undergone from the lowest level to the current one.*
Type:	*String*
Values:	*See structure*
Format:	*Complex*

Structure:	`<DATA_PROCESSING_HISTORY>`
	`<MODIFICATION>`
	`<ID>1</ID>`
	`<DATE_TIME_OF_MODIFICATION>`
	`YYYY-MM-DDThh:mmTZD </DATE_TIME_OF_MODIFICATION>`
	`<CURRENT_LEVEL>Current data product`
	`level</CURRENT_LEVEL>`
	`<PRECEEDING_LEVEL>Preceding data product`
	`level</PRECEEDING_LEVEL>`
	`<PROCESSING_TECHNIQUE>Indicate`
	`processing technique which was used to`
	`modify data from preceding to current data`
	`product</PROCESSING_TECHNIQUE>`
	`<AUTHOR>First_Name Last_Name of author`
	`that made modification</AUTHOR>`
	`<MESSAGE>Detailed description of the`
	`modification that has been made to data</MESSAGE>`
	`</MODIFICATION>`
	`...`
	`</DATA_PROCESSING_HISTORY>`

Mandatory:	*No, if data_processing_history is provided, otherwise yes*
Example:	

6.4 LIDACO

LIDACO (**LI**dar **DA**ta **CO**nverter) is a library and executable that enables a modular writing of FAIR lidar data converters for various types of lidars. The converter is based on the previously described data structure and format.

Currently, LIDACO supports conversion of data recorded by:

- Galion
- WLS70
- Windcube v1

- Windcube v2
- Windcube 100s, 200s and 400s
- long-range WindScanner
- ZephIR300

The converter and detail description on how to use it together with practical examples can be found at the
e-WindLidar public GitHub repository: https://github.com/e-WindLidar/Lidaco/

Chapter 7

Making data reusable

To make data reusable we envisage to provide a reference document, to which we refer as to measurement campaign and data report, that is intended to be consulted for information on the data, measurement campaign in which the data was created, instruments which were used in the experiment and personnel that was involved in the data creation process. The reference document is intended to be continuously updated with the latest information, thus its versioning is essential. The reference document is to a large extent based on the Perdigao paper Vasiljević et al. (2017). It consists of the following sections:

- Objective of measurements
- Site selection and description
- Experiment layout
- Lidar configuration
- Lidar calibration
- Data overview, data highlights and data usage
- Review of additional documentation
- Contributors

The section “Objective of measurements” is intended to describe in sufficient details the primary purpose of the measurements, basically answering the question “Why have we done the measurement campaign?”. The text should be understandable to scientific, non-scientific, technical and non-technical personnel. It should be written clearly and it does not necessarily need to conform to the rigor of the scientific writing. The same goes for other sections of the reference document.

The section “Site selection and description” should indicate why the site where the measurement took place was selected. It is essential to describe well the site characteristics. This includes terrain type, orography features, roughness classification, existing objects and infrastructures at the site, and if possible wind conditions (wind rose, turbulence statistics, and similar). The section should be accompanied with a graphical representation of the site and wind condition.

The section "Experiment layout" describes which lidars were used, where and how they were positioned. Any additional other additional instrumentation such as mast based sensors should be described in this section. It can also describe the power and network layout and access roads to instruments. This section, if possible, should indicate the constraints imposed by the site since this will indicate the rationale behind the selection of the lidar locations.

The section "Lidar configuration" provides details how lidars were configured during the measurement campaign. This includes description of the scanning strategies, probe length, measurement rate, and similar. One should consider this section of the reference report as an extension and elaborate description of the lidar specific metadata. Therefore, links to the metadata should be made here.

The section "Lidar calibration" report deployment and calibration procedural steps taken prior and during the measurement campaign (e.g., how leveling and orientation of the lidar were made) as well as the decommissioning and post-calibration procedural steps taken at the end of the campaign (e.g., post-campaign LOS calibration). This section should elaborate on the values metadata attributes related to the accuracy have. The detailed calculation of the lidar accuracy should be provided if possible.

The section "Data overview, data highlights and data usage" indicates data collection and archiving procedures, as well it provides a reasonably detailed overview of the collected data (e.g., a total amount of measurements prior and after filtering, number of 10-minute periods, etc.). Also, details where and how data can be acquired should be stated in this section. Besides the lidar data, if geographical data (terrain maps, roughness maps, point clouds, etc.) are available data creator should report them in this section. The same goes for the data acquired by other instruments (e.g., masts). This section should be accompanied by a number of observational highlights, which purpose is to make data users more motivated to explore the collected datasets. Since good datasets are used for many years following the completion of the campaign, it should be an imperative for data creators to record the usage of their data in this section of the report. This will be of help of any new data users to understand for what purposes and in what studies the data has been already in use.

The section "Review of additional documentation" should provide a list of additional documents, literature, presentation, web links and any other material that can be useful for data users and help them in their data exploration.

The section "Contributors" contains details on the personnel that contributed to the data creation. For the sake of participants visibility, it is recommended that range of tasks each contributor made is clearly outlined (e.g., contributors A designed experiment, monitored campaign, contributor B installed and calibrated lidars, etc.).

Examples of the filled templates can be found in Appendix A of this report.

Chapter 8

Conclusion

With the work presented in this communication, we initiated the alignment of efforts of lidar groups (to start with) at the European level in the data segment of the domain. In this work, we were inspired by the FAIR principles, and thus we presented an approach how to apply them to a specific data type. As result, we created guidelines and workflow for making lidar data FAIR. The guidelines and workflow can be used to make other domains data FAIR, thus they are universal. To our knowledge, this is the first application of the FAIR principles in wind energy sector and we hope that it will serve as a good example for other domains to apply the FAIR principles. The application of the FAIR principles led to the development of a generic lidar data format. This creates the basis for the development of community-based data processing tools since there is a standard approach in recording data. Also, we proposed a template for reporting data which will improve their reuse. Furthermore, we provided data conversion tools for several lidar types and examples of the converted datasets which are provided through the public Github repository https://github.com/e-WindLidar/Lidaco.

Appendix A

Reference documents examples

A.1 Kassel-2016

In the following a brief overview over the Kassel-2016 experiment is given as an excerpt from the detailed documentation: NEWA Forested Hill Experiment Kassel - Experiment Documentation, technical report IWES-KS-2017-214-V2. For more detailed information the reader is referred to that report.

A.1.1 Objective

Centred around the existing 200 m mast of Fraunhofer IWES at Rödeser Berg near Kassel an experiment on patchy forest over hilly terrain was conducted in 2016 and 2017. The NEWA forested hill experiment Kassel aims at characterizing the flow over a forested hill in a patchy landscape in relevant heights of modern wind turbines. It provides a unique dataset for model validation in this terrain.

A.1.2 Site selection and description

The NEWA forested hill experiment Kassel was conducted around the existing 200 m tall met mast of Fraunhofer IWES at Rödeser Berg. The Rödeser Berg is a hill 379 m above mean sea level in central Germany. The mast is located at the southwestern edge of a clearing (approx. 280 m north to south and 200 m east to west) on the ridge of the forested hill which stretches from approximately SSE to NNW. The closer surroundings of the mast are characterized by forest of varying heights and several clearings. The distance, up to which the forest stretches, strongly varies with direction. In the direction NNW the forest extends about 5.8 km, while in ENE the forest edge is already reached within approximately 400 m from the mast. The orography of the hill also varies strongly with direction. In general, the terrain is hilly and undulated. Towards NNW a hilly ridge extends for about 5.8 km. The wider surroundings consist of a patchy landscape of mainly agricultural land use, forest and some settlements. The immediate surroundings of the forested hill are mainly characterised by open agricultural areas. In the east and the west these are bordered by forested hills. In general, the terrain surrounding Rödeser Berg is very heterogeneous (see Pauscher et al. (2016), p. 2.)

A.1.3 Layout

The NEWA forested hill experiment Kassel aims at characterizing the flow over a forested hill in a patchy landscape in relevant heights of modern wind turbines. The main objective of the experiment is to provide a unique dataset for the validation of flow simulations in this terrain. For this purpose Rödeser Berg - a typical wind farm site in central Germany - has been chosen as a reference case for this kind of flow situation. The measurement campaign is focussed on the flow above and around the Rödeser Berg. The hill is aligned orthogonal to the main wind directions (southwest and northeast). The experiment consisted of a 3 month intensive campaign (from October 2016 to January 2017) and a 1 year long-term measurement campaign. The 1 year long-term campaign with two tall masts started in parallel to the intensive campaign. Both meteorological masts with heights of 200 m and 140 m were equipped with sonic and cup anemometers at multiple levels. The main focus of the intensive measurement campaign was the development of the flow over the ridge of the forested hill in the prevailing wind direction. A 5.5 km long transect along the main wind direction at 217° (counted clockwise against north) has been chosen as the flow line of main interest. The transect is split into two parts: upwind and downwind of the hill. The transect was probed with a dense array of instrumentation. The inflow conditions were determined with a 140 m tall mast, which also marks the starts of the transect. This mast was equipped with sonic and cup anemometers at 9 heights to allow for the characterization of wind and turbulence conditions. The 200 m tall met mast equipped with sonic anemometers at 9 height levels and a dense array of cup anemometers measured the vertical wind profile at the top of Rödeser Berg. The end of the transect was marked by a lidar wind profiler.

A.1.4 Configuration

In combination with the two tall masts, remote sensing devices and in particular multi-lidar measurements formed the backbone of the experiment. Two sets of synchronized long-range Doppler scanning wind lidars (i.e., the long-range WindScanner systems Vasiljević et al. (2016)) were used to create several virtual masts in step stare scanning mode. The location of the 140 m mast marked the starting point of the transect. The vertical wind profile of the 140 m mast was extended using a long-range lidar profiler (Windcube WLS 70) to heights of several hundred meters/few kilometres. This allows for the characterization of the flow aloft. The virtual masts were placed along the line between the 140 m crossing the 200 m tall mast and extending about 2 km behind the ridge of the hill. For each flow line (upwind and downwind of the hill), two synchronized WindScanners were used. The measurement heights of the virtual masts were set at 60 m (minimum realistic tip height above ground in forested areas) and at 135 m a. g. l. (hub height of the wind turbines on top of Rödeser Berg). Additional sampling points along the transect were provided by 2 wind profile lidars to support the virtual met masts and to provide continuous information on the wind conditions at the end of the flow line. Using the plan position indicator (PPI) mode, additional WindScanners measured the flow in front of the hill. The PPI overlay provides insights into the spatial distribution of the flow over the hill. As the number of vertical wind profilers (4 lidars and 2 sodars) and virtual masts was limited, additional information on the wind profile along the main stream line (transect along main wind direction) was desirable. Therefore, additional WindScanners carried out a range height indicator (RHI) scan from the start of the transect (location of the 140 m mast) and another one from the 200 m mast to the end of the flow line. 4 additional sites for wind profilers were selected in such a way that they could measure the incoming wind from other wind directions than the main wind direction. All measurement devices were placed and configured in a suitable way to assess the flow along the transect as well as in front (southwest, upwind) and behind (northeast, downwind) the hill (inflow and outflow) and in the wider surroundings. From southwest to northeast the flow along the transect is measured by (IDs of the measurement devices in parentheses):

1. The 140 m tall met mast to measure the inflow conditions (MM140)
2. A lidar profiler for great heights next to the 140 m mast (WP6)
3. A standard lidar profiler in the slope of the hill (WP3)
4. The 200 m tall met mast on top of the hill (MM200)
5. A standard lidar profiler in the lee of the hill (WP1)
6. A standard sodar profiler on a subsequent hill (WP5)

Additionally the following measurement devices have been used to measure the flow surrounding the hill (IDs of the measurement devices in parentheses):

- 2 lidar scanners as synchronized multi-lidar systems to measure the flow along the transects at multiple points upwind the hill (WS4 and WS5)
- 2 lidar scanners as synchronized multi lidar systems to measure the flow along the transects at multiple points downwind the hill (WS7 and WS8)
- 1 lidar scanner to perform RHI scans upwind the hill (WS1)
- 2 lidar scanners to perform RHI scans downwind the hill (WS2 and WS6)
- 2 lidar scanners to perform PPI scans upwind the hill (WS3 and WS9)
- 1 standard sodar profiler to measure the wind profile west of the hill (WP2)
- 1 standard lidar profiler to measure the wind profile south west of the hill (WP4)

In total 17 measurement systems have been used: 9 scanning lidars, 6 lidar/sodar vertical wind profilers, 2 met masts. The 2 masts have been measuring for one year in parallel (long-term measurements).

A.1.5 Data

Examples of acquired data and Lidaco config files are available at the e-WindLidar public GitHub repository:: https://github.com/e-WindLidar/Lidaco/tree/master/samples/Kassel_Experiment

A.1.6 Documentation

A detailed documentation of the Kassel-2016 experiment can be found in NEWA Forested Hill Experiment Kassel - Experiment Documentation, technical report IWES-KS-2017-214-V2.

A.1.7 Contributors

Fraunhofer IWES Kassel has conducted the measurements in corporation with DTU Wind Energy. Furthermore ForWind, Enercon, Innogy and Plankon/InnoVent have provided wind measurement equipment. The following persons and institutions have actively contributed to the success of the experiment:

- **DTU Wind Energy**: Michael Courtney, Per Hansen, Guillaume Lea, Søren William Lund, Jakob Mann, Robert Menke, Kristoffer Schrøder and Nikola Vasiljević
- **ForWind/Oldenburg University**: Joerge Schneemann and Stephan Voss
- **Enercon**: Michael Brüdgam and Lorenz Hutzler
- **innogy**: Meike Bilstein and Anthony Clarke Plankon
- **innoVent**: Dirk Ihmels and Roman Wagner vom Berg
- **Ge:Net**: Stefan Dümke and Sascha Engelaar
- **Telecon**:Lars Laurin

The following people and institutions have provided access to electricity and/or measurement sites: Arnd Gerhardt, Alfred Muth, Otto Elsner von der Malsburg, Deutsche Bahn, Matthias Schminke, EAM, Energie Waldeck-Frankenberg, E.ON, Maschinenring Kassel e. V., Stadtwerke Wolfhagen. Measurement team Fraunhofer IWES Kassel: Doron Callies, Richard Döpfer, Tobias Klaas, Alexander Kratzke, Sebastian Mehnert, Paul Kühn, Klaus Otto, Lukas Pauscher

A.2 Skipheia

A.2.1 Objective

The measurement campaign was performed as a validation study of a Windcube v2 lidar prior to the Valsneset campaign, where the lidar would be used as a standalone unit. The 10-minute mean horizontal wind speed measurements are compared to 2D ultrasonic anemometer (Gill WindObserver II) measurements at 40, 70 and 100 meters above ground level from a 100 meter met-mast.

A.2.2 Site selection and description

The measurements were performed at the Skipheia met-station at the island Frøya, on the Mid-Norwegian coast. A map showing the location of the site near the village Titran is shown in Figure A.1. The site consists of two 100 meter met-masts and one 45 meter met-mast in a triangular configuration. The 100 meter masts, equipped with 2D ultrasonic anemometers, are suitable for validating the lower third of the vertical range of the Windcube. The site is at an exposed coastal location, often subjected to harsh weather conditions. This is also reflected in the surrounding ground cover which is dominated by grass, marshland, heather and bare rock. The surrounding terrain is not completely flat, but consists of small rolling hills with a height difference of 20 meters. The distance to the shoreline varies between 0.3 and 30 kilometers and the height above mean sea level is 20 meters. In figure 1 the height contours and the undisturbed measurement sector is shown. The wind rose measured by the lidar during the campaign is shown in Figure A.2.

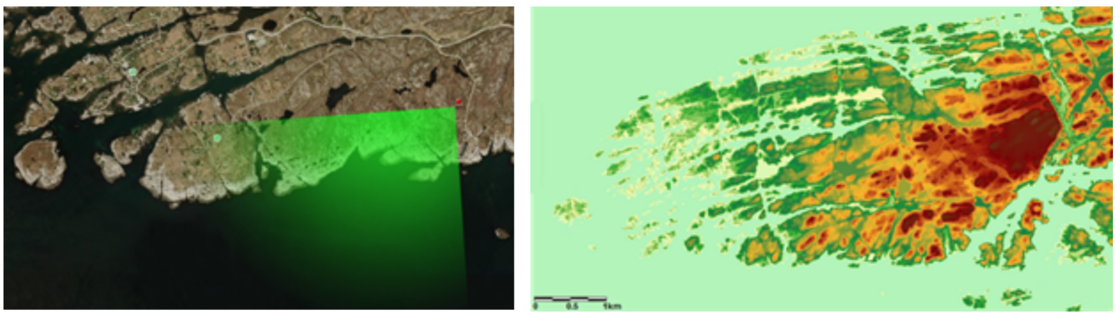


Figure A.1: Left - Undisturbed measurement sector, Right - Contour map showing the surrounding terrain. The height range is 40 meters.

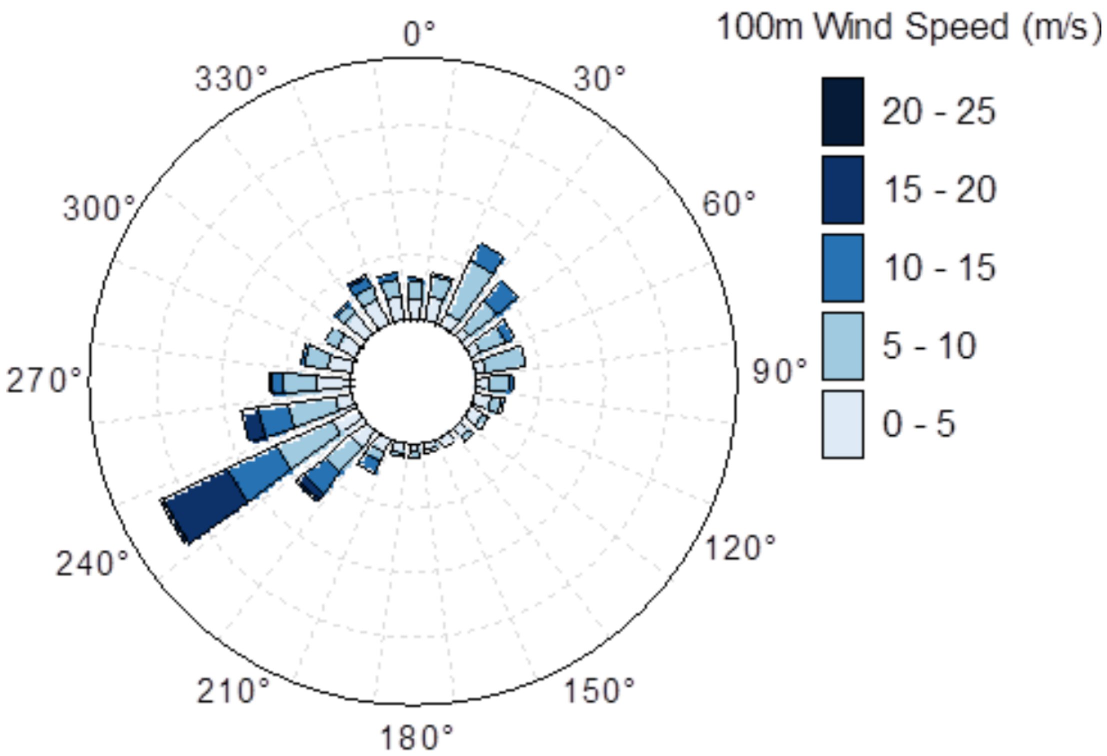


Figure A.2: Wind rose from the 30 days lidar campaign

A.2.3 Layout

A single Windcube v2 (WLS7-171) from Leosphere was used in this campaign. The lidar was placed on the ground at a distance of 10 meters from the base of a 100 m high met-mast. The azimuth angle towards the mast is 41°. The lidar was positioned and oriented such that neither of the lidar beams would interfere with the mast structures. The direction offset is -12.5° relative to true north. The 45° sector used in the comparison with boom-mounted Gill WindObserver II ultrasonic anemometers is coincident with the dominating wind direction as seen in Figure A.3.

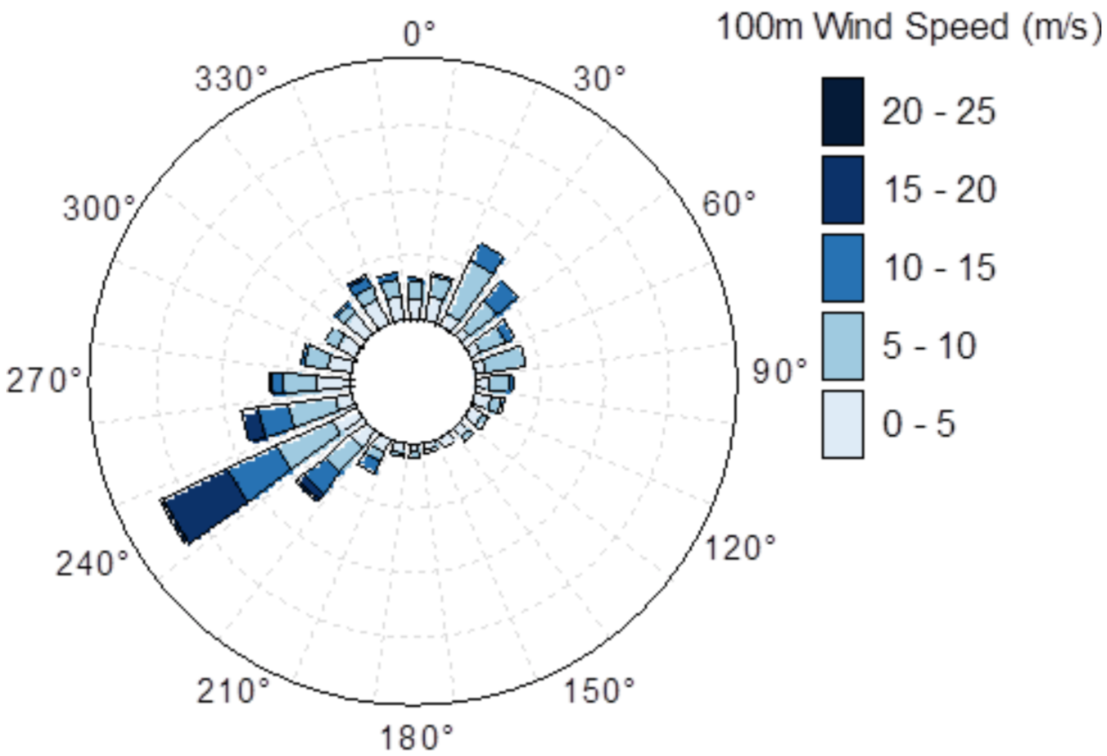


Figure A.3: Wind rose from the 30 days lidar campaign

A.2.4 Configuration

The lidar was used in a standard factory configuration from Leosphere. The Windcube v2 is a pulsed lidar with four inclined beams and one vertical beam, using a Doppler Beam Swinging technique (DBS) to calculate the horizontal wind speed. The scan angle is 28° relative to vertical . It emits 20000 pulses per LOS measurement at a pulse repetition rate of 30 kHz. The pulse duration is 175 ns. The accumulation time is not constant. It is follows a cycle of 0.72 seconds for 4 of the beams and 0.96 seconds for the fifth beam. Since the fifth beam is a vertical beam, an independent horizontal wind vector can be calculated every 3.84 seconds. The configured measurement heights are 40, 50, 60 70, 80, 90, 100, 120, 150 and 200 meters above ground level.

A.2.5 Calibration

The lidar was oriented at an angle of -12.5° relative to true north, as measured by the lidars internal compass. The unit was placed on solid rock and leveled manually using the internal inclinometer. No further calibration or validation was made prior to the measurements as this was the aim of the measurement campaign. The 10-minute average wind speed accuracy, given as min-max bias versus reference anemometry, is given by Leosphere as 0.1 m/s.

A.2.6 Data

Examples of acquired data and Lidaco config files are available at the e-WindLidar public GitHub repository: https://github.com/e-WindLidar/Lidaco/tree/master/samples/Skipheia

A.2.7 Documentation

The measurement campaign and the main results are presented in the following unpublished paper: Bardal L.M., Comparison of lidar and met-mast wind measurements. 9th EAWE PhD seminar on wind energy in Europe, Uppsala University Campus Gotland; Visby, Gotland. 2013-09-18 - 2013-09-20

A.2.8 Contributors

The installation of the lidar and data collection was performed by Lars Morten Bardal (NTNU). Data conversion was performed by Felix Kelberlau and Lars Morten Bardal, using a conversion tool made by Tobias Klaas (Fraunhofer IWES).

Bibliography

Berg, J., Vasiljević, N., Kelly, M., Lea, G., and Courtney, M. (2015). Addressing Spatial Variability of Surface-Layer Wind with Long-Range WindScanners. *Journal of Atmospheric and Oceanic Technology*, 32(3):518–527.

Courtney, M., Wagner, R., and Lindelöw, P. (2008). Testing and comparison of lidars for profile and turbulence measurements in wind energy. *IOP Conference Series: Earth and Environmental Science*, 1:U172–U185.

Kameyama, S., Sakimura, T., Watanabe, Y., Ando, T., Asaka, K., Tanaka, H., Yanagisawa, T., Hirano, Y., and Inokuchi, H. (2012). Wind sensing demonstration of more than 30km measurable range with a 1.5 μm coherent doppler lidar which has the laser amplifier using er,yb:glass planar waveguide. *Proc.SPIE*, 8526:8526 – 8526 – 6.

Pauscher, L., Vasiljević, N., Callies, D., Lea, G., Mann, J., Klaas, T., Hieronimus, J., Gottschall, J., Schwesig, A., Kühn, M., and Courtney, M. (2016). An Inter-Comparison Study of Multi- and DBS Lidar Measurements in Complex Terrain. *Remote Sensing*, 8(9):782.

Rodrigues, C. V., Palma, J., Vasiljević, N., Courtney, M., and Mann, J. (2016). Coupled simulations and comparison with multi-lidar measurements of the wind flow over a double-ridge. *Journal of Physics: Conference Series*, 753(3):032025.

Rumbaugh, J., Jacobson, I., and Booch, G. (2004). *Unified Modeling Language Reference Manual, The (2Nd Edition)*. Pearson Higher Education.

Sempreviva, A. M., Verelst, D. R., Giebel, G., Mikkelsen, L. P., Bak, C., Danielsen, H. K., Vasiljevic, N., Vesth, A., Anderson, M., Barth, S., Rodrigo, J. S., Gancarski, P., Reigstad, T. I., Bolstad, H. C., Wagenaar, J. W., and Hermans, K. (2017). Taxonomy and meta data for wind energy R&D. Technical report, IRP-Wind, Brussels, Belgium.

Sobolev, V. S. and Timokhin, S. A. (2014). Maximum likelihood estimates of signal parameters of laser doppler systems. *Optoelectronics, Instrumentation and Data Processing*, 50(6):566–576.

Vasiljević, N. (2014). *A time-space synchronization of coherent Doppler scanning lidars for 3D measurements of wind fields*. PhD thesis, Technical University of Denmark (DTU). PhD-0027.

Vasiljević, N., L. M. Palma, J. M., Angelou, N., Carlos Matos, J., Menke, R., Lea, G., Mann, J., Courtney, M., Frölen Ribeiro, L., and M. G. C. Gomes, V. M. (2017).

Perdigão 2015: methodology for atmospheric multi-doppler lidar experiments. *Atmospheric Measurement Techniques*, 10(9):3463–3483.

Vasiljević, N., Lea, G., Courtney, M., Cariou, J.-P., Mann, J., and Mikkelsen, T. (2016). Long-Range WindScanner System. *Remote Sensing*, 8(11):896.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., and Dumon, O. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(160018).