The rise of big lidar datasets and need for lidar data standardization, contextualization and dissemination

N. Vasiljević*, A. Vignaroli*, C. Hasager*, L. Pauscher, T. Klaas**, J.C. Lopes***, H.C. Bolstad**** and L.M. Bardal*******
*DTU Wind Energy, Roskilde, Denmark, niva@dtu.dk
**Fraunhofer IWES, Kassel, Germany, lukas.pauscher@iwes.fraunhofer.de,
*** UPORTO, Porto, Portugal, jlopes@fe.up.pt
**** SINTEF, Trondheim, Norway, hans.christian.bolstad@sintef.no
***** NTNU, Trondheim, Norway, lars.m.bardal@ntnu.no

Abstract

Lidars are routinely operated in the wind energy sector with the aim of creating high-quality datasets for atmospheric flow studies. Since lidars are managed by different groups, virtually each dataset is stored and described differently. This results in inefficient work of data users, as they need to get familiar with various data formats to handle datasets from different data providers. To tackle this problem, several European lidar groups initiated work towards the lidar data standardization, contextualization and dissemination. In this talk, we present preliminary results of this work, which has been carried out under the e-WindLidar project.

Keywords: *lidar data*, *open data*, *e-Science*

Introduction

Lidars have the capability of acquiring high-quality datasets of atmospheric flows which are of interest for the wind energy sector [1]. A significant number of unprecedented lidar datasets has been collected throughout several large national (e.g., RUNE campaign [2]) and international experiments (e.g., Perdigão-2015 [1], Perdigão-2017 [3,4], Kassel-2016 [4]) that are intended to be publically available in near future. As the lidars in these experiments were operated by different lidar groups a need emerged for consensus on how the acquired datasets should be stored, described and disseminated in order to simplify the work of data users while improving the visibility of data creators. Several WindScanner.eu [5] partners initiate the work towards the lidar data standardization, contextualization and dissemination under the e-WindLidar project. This project is a part of the Open Data (OD) initiative which is carried out within the Joint Programme on Wind Energy (IRPWind) of the European Energy Research Alliance (EERA).

The partners set a target to build and operate Open Access (OA) community-based web platform "e-WindLidar" for lidar data and data analysis tools dissemination. In this talk, we focus on the fundamental work that has been carried out prior the platform development. This work includes the classification of lidar data products, development of comprehensive lidar data model for the most frequently stored product and selection and adaptation of data storing format in accordance with the model.

Approach

The lidar data have been classified into five distinct data products, which correspond to the levels in lidar data processing. These products are: backscatter signal (Level 0), Doppler spectra (Level 1), Line-Of-Sight (LOS) wind speed (Level 2), wind vectors (Level 3) and flow parameters/statistics (Level 4). Levels are interconnected by means of different data processing techniques.

Since Level 2 product is a product that is usually produced and stored in real-time by lidars, for this product we developed a comprehensive lidar data model and standardized how the product should be stored. While building the model, we identified three types of information that impact how lidar data are stored and what metadata are reported. These are: details on the instrument used to produce data, measurement configuration of the used instrument and information on the data themselves (e.g., to what kind of quality control the data were subject to). The model is generic; thus it can describe datasets of lidars with a fixed or flexible scanning geometry installed on a static or moving platform (e.g., scanning or profiling lidars installed on the ground, floating platform, vessel, etc.).

We selected NetCDF file format for storing lidar data and metadata. A detailed list of attributes, which represent lidar specific metadata, and variables to accommodate various deployments and configurations of lidars was created. We proposed the structure of the NetCDF file which identically to the model is generic, thus it is independent of the lidar technology, measurement configuration or lidar deployment type. Lidar specific metadata are intended to help data users in their data analysis making them (to large extent) independent from data creators. Nevertheless, these metadata are also intended to provide traceability of the instrument, measurement configuration, and data, transparency of the work that has been done to create the data and the visibility of data creators.

Conclusions

We laid down foundations for the development of the e-WindLidar web platform. Namely, we classified lidar data to distinct products, thus creating the lidar data taxonomy, modeled the most frequently produced lidar data product (i.e., Level 2 – LOS speed) and standardized how this product should be stored. The lidar data taxonomy indicates the complexity of the lidar data analysis and levels of lidar data refinement. It is an important tool for assessing the required workload of data creators prior measurement campaigns but also the workload of the data users during the data exploration phase. The Level 2 data model and corresponding file storing format represent the basis for the development of lidar data conversion tools. Since the lidar data format is generic the future tools that that will handle analysis of this new format can be reusable. Moreover, the standardized lidar data format represents fundamentals for the development of community-based tools, both for data conversion and data analysis.

References

1. N. Vasiljević, J. M. L. M. Palma, N. Angelou, J. C. Matos, R. Menke, G. Lea, J. Mann, M. Courtney, L. F. Ribeiro, and V. M. M. G. C. Gomes, Perdigão 2015: methodology for atmospheric multi-Doppler lidar experiments, AMT, 2017
2. Floors, R., Peña, A., Lea, G., Vasiljević, N., Simon, E., and 105 Courtney, M.: The RUNE Experiment—A Database of Remote-Sensing Observations of Near-Shore Winds, Remote Sensing, 8, 2016.
3. Witze, A.: World's largest wind-mapping project spins up in Portugal, Nature, 542, 282–283, http://www.nature.com/news/world-s-largest-wind-mapping-project-spins-up-in-portugal-1 , 21481, 2017.
4. Mann, J., Angelou, N., Arnqvist, J., Callies, D., Cantero, E., Arroyo, R. C., Courtney, M., Cuxart, J., Dellwik, E., Gottschall, J., Ivanell, S., Kühn, P., Lea, G., Matos, J. C., Palma, J. M. L. M., Pauscher, L., Peña, A., Rodrigues, C. V., Rodrigo, J. S., Söderberg, S., Vasiljević, N.: Complex terrain experiments in 75 the New European Wind Atlas, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 375,
5. http://www.windscanner.eu