|  |
| --- |
| Vestas Climate Library  A logo of a company  Description automatically generatedDescription of the available data |

A picture containing drawing

Description automatically generated

Contents

[Objective 3](#_Toc151648948)

[VCL data access 3](#_Toc151648949)

[Description of the VCL data 3](#_Toc151648950)

## Objective

The objective of this short document is to describe the extent and the characteristics of the Vestas Climate Library (VCL) data available to the hackathon.

## VCL data access

The VCL data will be available either as static files in the hackathon data folder, or through an API directly accessing the main VCL data system in Vestas.

The API will be accessible to the hackathon participants via prior registration and approval, and using the credentials valid during the hackathon. The access may be available for a limited period before and after the event, per individual agreement.

## Description of the VCL data

The VCL data has been generated by a numerical weather prediction (NWP) model simulation.

A graph of a graph

Description automatically generated with medium confidenceA green squares with black text

Description automatically generatedA map of the state of wyoming

Description automatically generatedA graph of a graph

Description automatically generated with medium confidence

Figure 1: Left: an illustration of the spatial representation of the VCL data. Center: a map of an aggregated quantity, in this case a proportion of time when the wind speed 100 meters above the surface is less than 4 m/s during daytime. Right: a time-series of a simulated and measured wind speed.

The weather variables like wind speed, wind direction, temperature, humidity, cloud density (water and ice), precipitation, turbulent kinetic energy, pressure, air density, sensible and latent heat flux, solar radiation, etc., are available hourly since the year 2010 until one month before the present, at a horizontal resolution of 3 kilometers, at any height above the surface. Certain offshore regions may not be covered with the high resolution data and a 15 km resolution will be available. For smaller areas, higher resolution data may be produced in advance, per individual agreement.

Two kinds of data will generally be available: time-series, and time-aggregated weather variables. The time-series data is useful for detailed analyses involving identification of regimes, correlations with other data sources, and similar. The aggregated data is typically used to provide area overviews of certain relevant conditions, for example average temperature, amount of rain, or frequency of occurrence of certain conditions met over an area. A few examples of these data types are shown below. It is expected that the hackathon participants will be free to choose the data type and extent. However, if a participant’s intended investigation requires more data than it may be accessible via the foreseen channels, different kinds of VCL data queries may be designed. That may even include invoking a specific user-defined function during the data aggregation.

### Example 1 – time-series

A screen shot of a computer screen

Description automatically generated

In this example, the data is delivered as a time-series of several variables at several heights above the ground. Any height between 20 m and about 800 m above the ground is electable; at every second model grid point (i.e. every 6 km in each direction) the upper limit is about 20 km. There are 22 available variables with the vertical dimension, and 45 available variables at either the surface, in a layer, or at a pre-defined height, i.e. the sea-surface temperature, the boundary layer height, or the temperature 2 m. The data frequency is 1 hour but can be up to 15 minutes if needed, and the period covered is from the year 2010 to recently, but can be extended in the past if needed. The variables, heights, and the time period can be chosen at the time of querying the VCL.

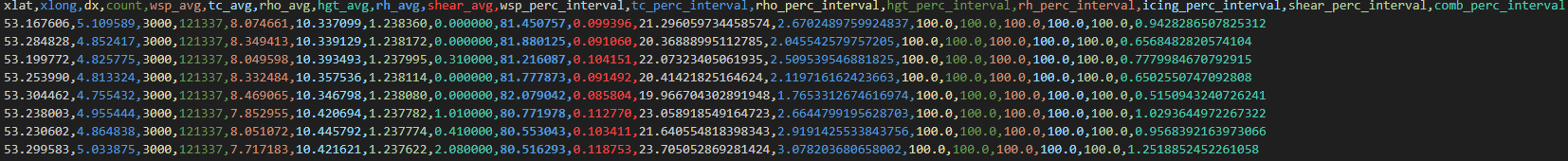
### Example 2 – aggregated values

A black screen with numbers

Description automatically generated

The data can also be aggregated as simple averages (the table above), or as a minimim/maximum value, or as a standard deviation, at any height above the surface. The variables, heights, and the time period for aggregation can be chosen at the time of querying the VCL.

### Example 3 – probability of occurrence



It is possible to present the climate data as a proportion of time when certain filtering criteria are met. In the table above, for every geographical location in a chosen area, the average values of several climate variables are listed, followed by the proportion of time when the wind speed and the air temperature are within a user-defined interval (in this case below 5 m/s and above 20 degrees C – a use case for calm loving sunbathing tourists), and at the end the combined probability that all the filtering criteria are met. The filtering variables are pre-defined but can be enabled or disabled; one height can be selected, and the time period for aggregation can be chosen at the time of querying the VCL.

### Example 4 – time-series over an area

A screen shot of a computer

Description automatically generated

When the time-series data is extracted over an area, the selected variables will be available for every point in the area for every hour of the period. This data format is in fact the most similar to how the simulation data itself is stored. Because the volume of data, resulting from this query is very large, it will always represent just a limited sample and shall be used as a template for deriving solutions which eventually would address much larger areas and longer time periods. The solution implementation which works on a sample of data subject to the volume limitation (80 million data items) shall eventually be programmed directly into the Databricks notebook, integrated into the VCL itself.