**Explanation to the wind park data set**

Which of the variables do you want us to consider to visualize in the pie for each turbine?

IS (Dr. Irene Schicker Geosphere Austria): I would suggest those that have highest importance but at least the first five or six variables (please with fixed colors per variable).  
  
I could see there are two air pressure  levels = 900, 925 given  for each turbine  and for these corresponding measurements. Does it mean that level =900 corresponds to the lower part of the movable turbine blade and level =925  to the upper part of the movable turbine blade?

IS: I would suggest to use the 925. It is still quite higher than the tip of a turbine blade but given that ERA5 is known to underestimate wind speed especially in the lowermost levels 925 is a good trade off (it corresponds roughly to 600m, depending on temperature etc.).  
  
Variable "height above ground" is constant.

IS: It gives you an indication of which height the model sees it. You can ignore it (but in general good to have it to have an idea of the data)  
  
Do you want to visualize the  relative importance  of causal influence of "height above gound" with respect to the remaining variables indifferent scenarios (high extreme wind, moderate wind, low extreme wind)?  
IS: We do not really need to consider that parameter, it is there just for completeness.  
Also variables longitude  17.048147, latitude   47.803306 seem  to be constant  over the years. If they differ, the difference is in orders of 0.01, so it looks like as a noise by measurements. Anyway, longitude and lattitude serve to identify the turbine, which are at the same time identifed by the number of each turbine, so these two  
variables are redundant.

IS: I gave you the information so that the metadata of the turbines is included as well (and to plot them yourselves on a map). The turbines of this wind park are in operation starting for quite some years now and they did not change location. It is also to be on the safe side in case you/we/whoever wants to add additional information to have more than only the id describing the data.   
  
I would definitely omit longitude and lattitude  and I would tend to omit also  height. I do not think the causal influence  of these three variables will significantly differ  with respect to extreme scenarios.

IS: These parameters make sense if you have more than one wind park. For now, with only one, you can omit them.  
  
Why did we get 42 files and not 21 (for each year from 2000 -2020 one)?

IS: This is indicated by the name and the type of variable: \*surface\* indicates that the parameters are derived from ERA5 single layer (=surface) including 100m wind speeds. Thus, the wind speed you find in the surface files are those corresponding to hub heights (in theory). The \*pressure\* ones contain other parameters (like cc- cloud cover) and wind speed values at the pressure levels above ground. And to me it makes sense to also include the pressure level wind speed as, see above, ERA5 has some issues with wind in the lower levels of the atmosphere.

There are in different files as the extraction procedure and some calculations differ and because the extraction of the pressure level data takes considerably longer than the surface based data.