The measure-distribute-predict (MDP) method

1 Background and scope

aj (anemos-jacob GmbH) has been using the Measure-Distribute-Predict (MDP) method to extrapolate wind speed data measured at examined sites to the long term since it was developed in house in the year 2003. However, no detailed explanation has been released so far. Due to a number of requests from third parties for details, the present document describes the reasoning behind the method and its concept. It further shows results from an application to demonstrate its power.

The present document only addresses the use for wind resource assessment in the wind energy sector.

The data to be extrapolated to the long term is generally called "on-site data" or "data of the examined site" in the present document. This is typically data collected on temporary met masts or with remote sensing instruments. The "long term reference data" or "reference data" are typically data measured on meteorological stations or calculated as part of re-analysis or mesoscale data sets.

2 Motivation, deficiencies of the MCP methods

The most common approach to scale measured on-site wind data to the long term is called Measure-Correlate-Predict (MCP). Its basic idea is to establish an empirical relationship between long term reference data and on-site data based on a concurrent data set stretching over a shorter period (typically one year or a few years). This relationship is then applied to the long term statistics (or time series) of the reference data in order to synthesise long term statistics (or virtual time series) at the examined site.

This basic principle is maintained for the MDP method, but it uses a different type of transfer function. It should be noted though that a large range of different approaches are currently called MCP. No consensus regarding important details is visible. The main difference between these variants is the type of transfer function. Thus, the MDP method adds yet another type of transfer function to this range of methods.

The fundamental difference of the MDP method compared to the common MCP approaches is how scatter between on-site data and reference data is considered and treated. This scatter is not attenuated; it is accepted and maintained instead. This needs to be explained and justified.

Scatter between on-site data and reference data is usually caused by a combination of

- Time lag due to the passage of weather events between the locations
- Local topographic influences
- Regional topographic and atmospheric influences
- Atmospheric stratification
- Stochastic processes in the atmosphere
- Differences between spot observations and averaged assumptions in the case of re-analysis or mesoscale reference data

Although the above aspects are generally known, some of them are illustrated in the following in order to explain the motivation of the MDP method.

2.1 Time lag, stochastic processes, local topographic influences

The following graph compares the 10 minute average wind speed data recorded at 100 m height above the ground on two met masts installed in flat terrain at about 9 km distance from each other along the prevailing wind direction.

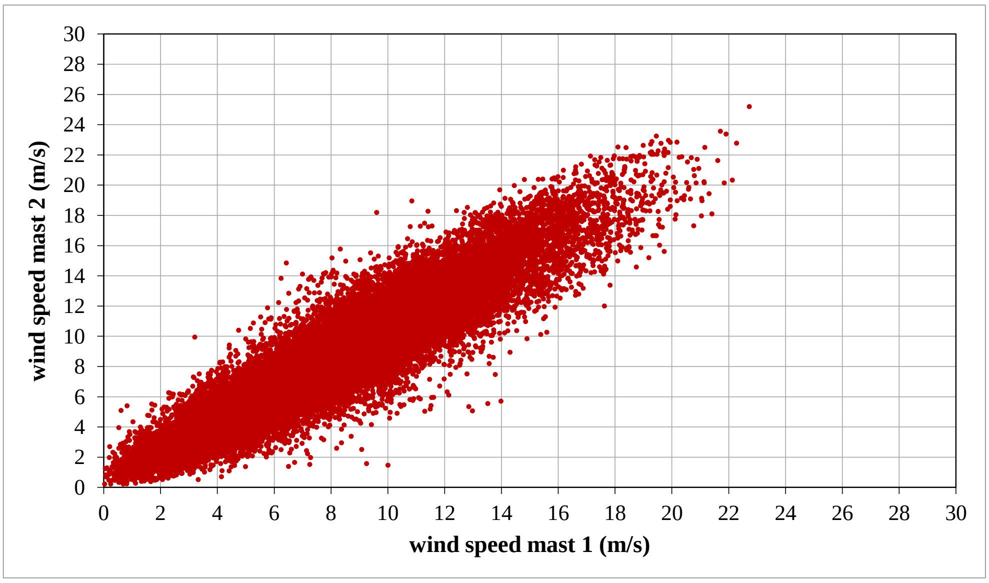

Figure 1: Comparison of concurrent 10 minute average wind speed data recorded at 100 m height with 9 km distance

Considering the simplicity of the terrain, the amount of scatter is enormous. Usually, the distance between the location of a reference data set (i.e. typically met station) and the considered site is larger and the local topographic influence is likely to be more relevant as well. Thus, an amount of scatter even bigger than in Figure 1 must be accepted as natural and it does not necessarily indicate that the reference data set is not representative for the examined site.

The common MCP procedures resolve the problem of scatter by calculating a compromise function such as a linear regression (or sometimes non-linear regression). Usually, for Figure 1 this would mean that all values of wind speed recorded on mast 2 when a specific value is observed on mast 1 are averaged. There are several statistical problems connected with this approach, most of which should not be discussed in the present document. Just two of them should be mentioned:

1) Such approach tends to reduce the bandwidth of values.
2) It neglects the fact that the energy contained in the wind does not depend linearly on the wind speed, i.e. the averaging process changes the energy contents contained in the data set.

The importance of these issues increases with increasing scatter. The scatter cannot be reduced for a given averaging time, since the locations of the reference data set and the investigated sites are fixed.

However, if the data are averaged over longer time intervals, the scatter can be reduced. For Figure 1, averaging to hourly means leads to the following picture:

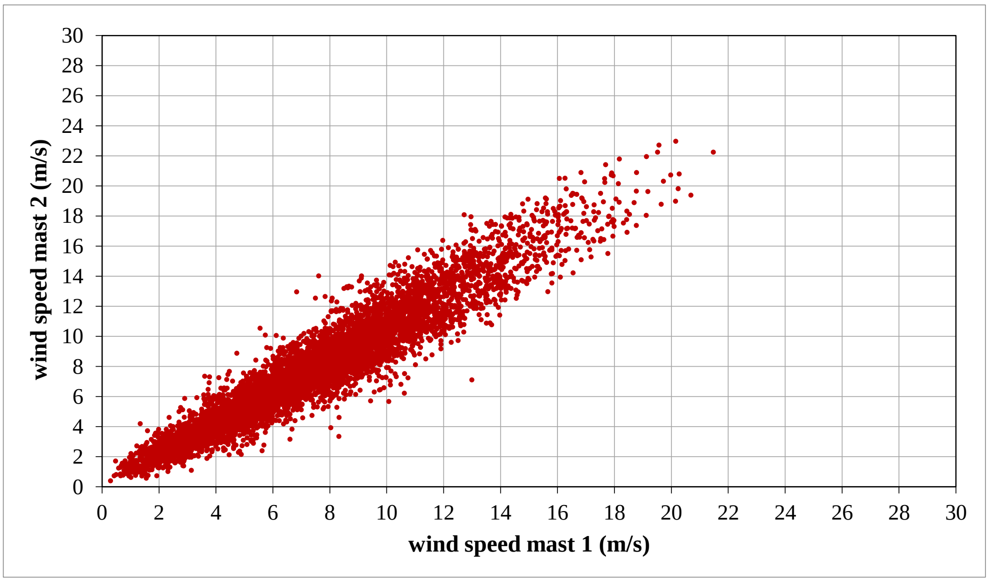

Figure 2: As for Figure 1, but using hourly averages

The scatter has visibly been reduced, but the coefficient of determination only increased from 0.88 to 0.91. At the same time, the maximum wind speeds appearing in the data sets were lowered and some error regarding the energy contents is made within each data point. Thus, the apparent increase in accuracy in the procedure and of representativeness between the data sets was achieved by losing information. This means in general that attenuating the scatter through statistical means may be linked with loss of accuracy.

2.2 Atmospheric stratification

The impact of stratification is relevant in many aspects of wind resource assessment, which is commonly under-estimated. This impact increases with increasing considered height ranges above the ground, i.e. generally with increasing hub heights.

The following graph compares wind speed data measured on two anemometers installed at 10 m height and 100 m height on the same mast.

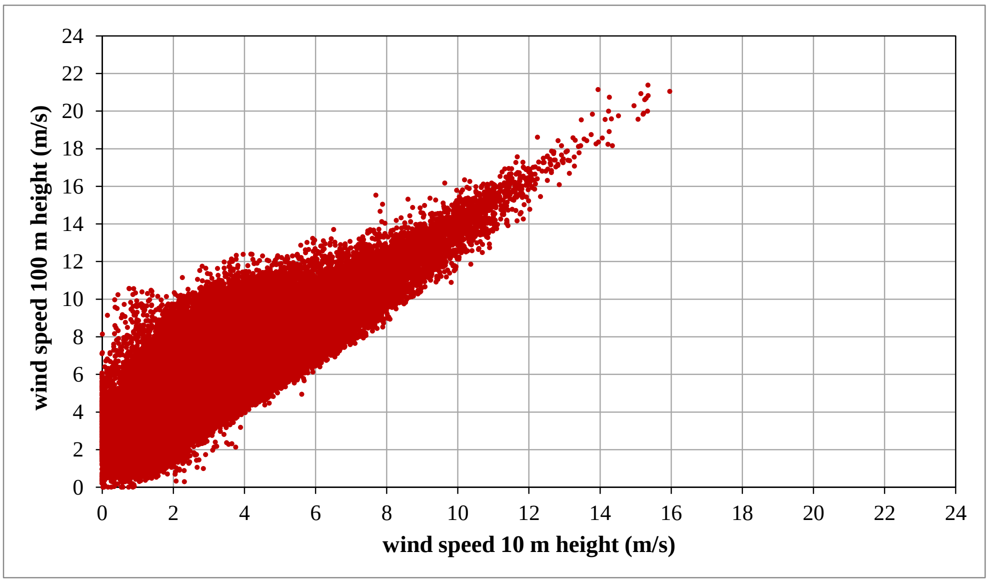

Figure 3: Comparison of concurrent 10 minute average wind speed data recorded at 10 m and 100 m height on the same met mast

The large scatter in Figure 3 is mainly due to atmospheric stratification. Turbulence causes only some additional scatter. The impact of the met mast on the measurement is very small in this case. It therefore hardly affects the picture.

It is re-iterated that common MCP methods involve averaging across all wind speeds observed in the analysed measurement for a given wind speed in the reference data set.

If in the above case, the 10 m height measurement was the reference data set and the 100 m height measurement the on-site data, it would be difficult to argue that the reference data set cannot be considered as a valid and representative reference data set for the on-site measurement. This would simply mean discarding large ranges of potential reference data sets in a very general sense. However, if the 10 m height measurement was a long term reference data set, any extrapolation method based on averaging of the on-site observations for a 10 m height wind speed would obviously cause high uncertainties.

This Figure is of interest because many reference data sets have been measured at or refer to 10 m height above the ground. Since Figure 3 has been established using wind data from the very same location, that the correlation between any more common reference data obtained at or for 10 m height in the region of interest and the wind speeds at 100 m height on this site is likely to be worse than that of Figure 3. It is also obvious that the problem becomes even more pronounced with higher measurement levels. However, if such patterns occur between reference data and on-site data in wind resource assessments, it is often argued by parties carrying out due diligence that the reference data are not sufficiently representative for the investigated site. Figure 3 shows that as long as the reference data have not been acquired at or for higher levels above the ground, this scatter cannot be avoided. It is an indication for missing representativeness only with respect to stratification, but not with respect to the location, measurement accuracy or other aspects. Clearly, in normal circumstances issues like location or measurement accuracy add to the scatter already caused by stratification.

Thus, any extrapolation procedure that shall deliver accurate results must allow for patterns as in Figure 3. It must try to account for these effects as much as possible. One could simply say: If the relationship between the reference data and the on-site data is far from linear, then linear regressions should not be used for the extrapolation. In the view of *aj*, this also applies to any further method that involves averaging if the relationship does not follow permanently more or less the same trend.

Clearly, if stratification is an issue as in the present example, averaging of the data over longer periods (several days or more) will often improve the pattern. For Figure 3, this is demonstrated and discussed in the following.

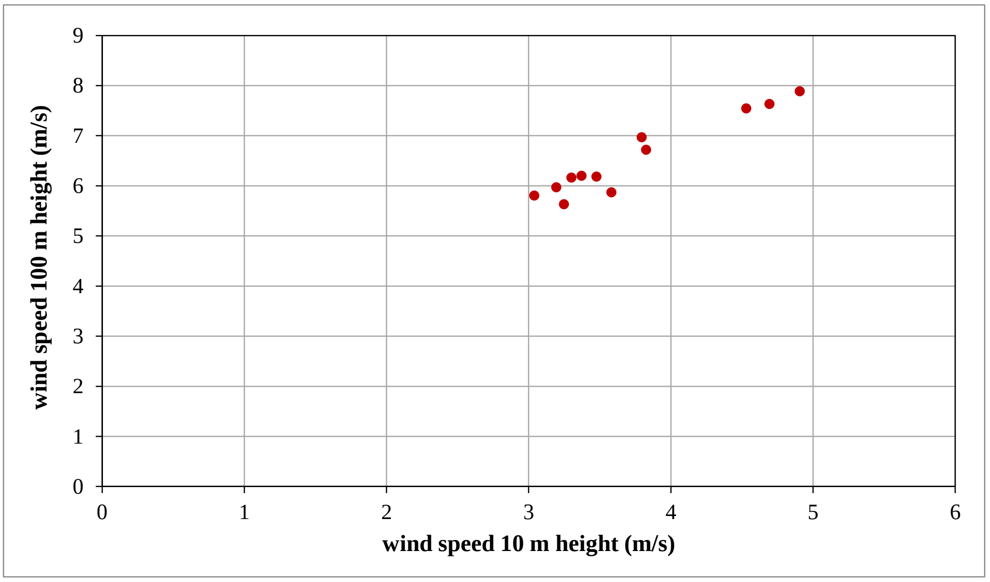

Figure 4: As for Figure 3, but based on monthly averages

The points deviating most from the general trend in Figure 4 relate to the months of August and September. Both were particularly sunny months in the present case, the average air temperature in September was above that of August. Thus, varying stratification still had an impact on the relationship between the wind speeds at 10 m and 100 m height even on a monthly basis.

Nevertheless, Figure 4 clearly shows that the impact of stratification can largely be attenuated through averaging. However, the comments made above on the impact of averaging wind speed data apply here even more. In particular, the information on the frequency distribution of the wind speed, hence the energy contents, gets lost with averaging over longer time periods. This becomes obvious from the following graph.

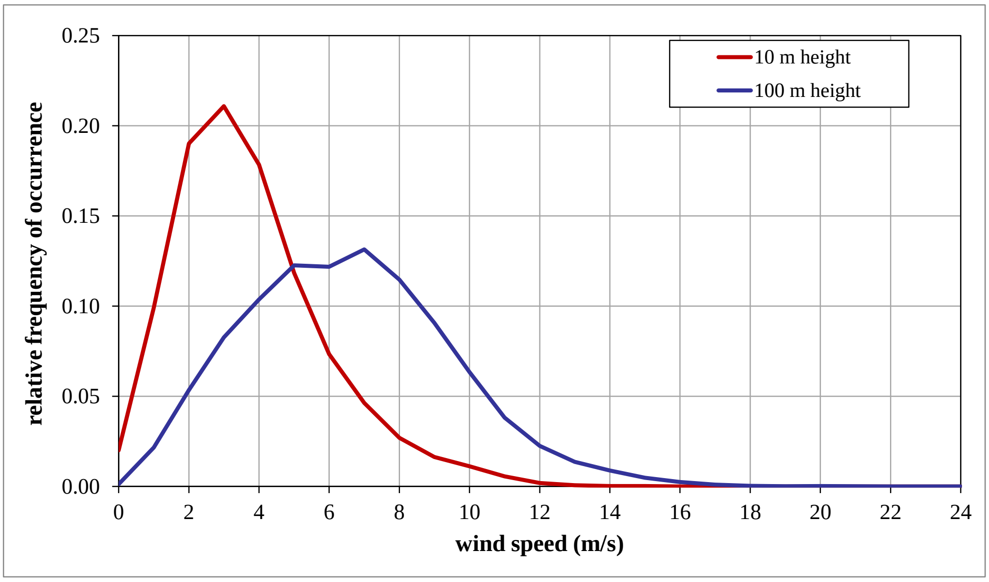

Figure 5: Frequency distributions of the wind speeds of Figure 3

The shapes of the frequency distributions of the wind speed recorded at different heights differ significantly. Consequently, the change of wind energy between the heights will not be proportional to the change of mean wind speed. Thus, simple scaling methods between the mean wind speeds will alter the energy contents. This also applies to linear or non-linear regressions performed over large scatter such as in Figure 3.

Again, differences in shape between frequency distributions such as in Figure 5 are sometimes used as an argument against the selection of reference data during due diligence processes. With large height ranges between reference data and on-site data, such differences may be unavoidable. Therefore, they will have to be accepted. Clearly, care must be taken in these cases regarding the extrapolation procedure. In particular, the resulting long term frequency distribution for the investigated site and height above the ground must be based appropriately on that measured there. It should not be a scaled up version of the frequency distribution of the reference data set.

2.3 Regional topographic and atmospheric influences

Obviously, if topographic or atmospheric influences or even a combination of both lead to differences in the wind field between the examined site and the location of the reference data, the situation becomes even more demanding for the extrapolation procedure.

In the following example, the examined site is located at higher altitude at a moderate distance from the Mediterranean Sea. A met station is located at just a few kilometres distance from the site, but much lower in a large valley. It was considered as reference data source.

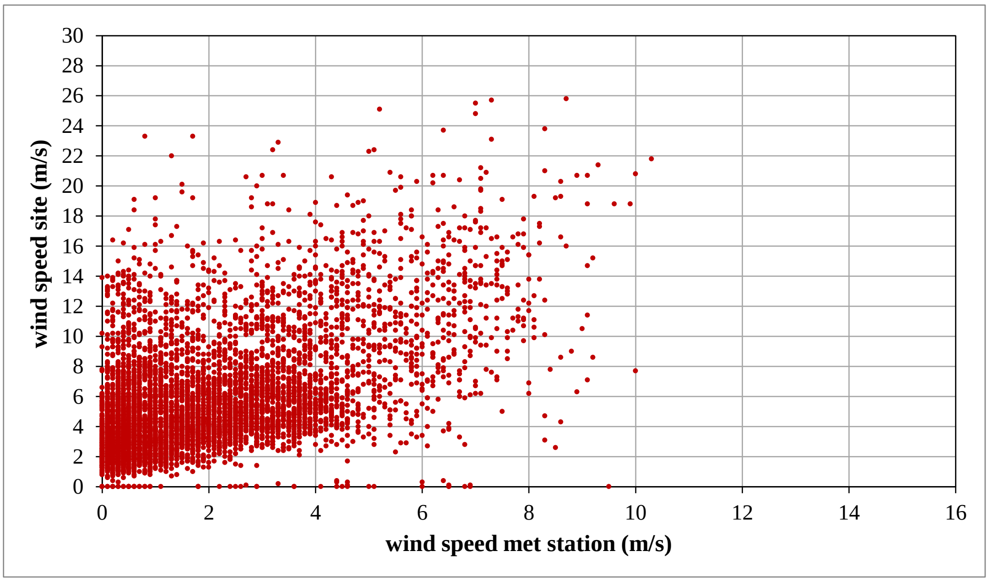

Figure 6: Concurrent hourly wind speeds of a site and a met station in the Mediterranean region

Figure 6 seems to suggest that no connection exists between the met station and the site. In such case, the met station could not serve as long term reference. However, a more detailed analysis showed that both wind fields are not independent from each other, but that their relationship is very complex. The comparison of the wind direction data provides an indication for this.

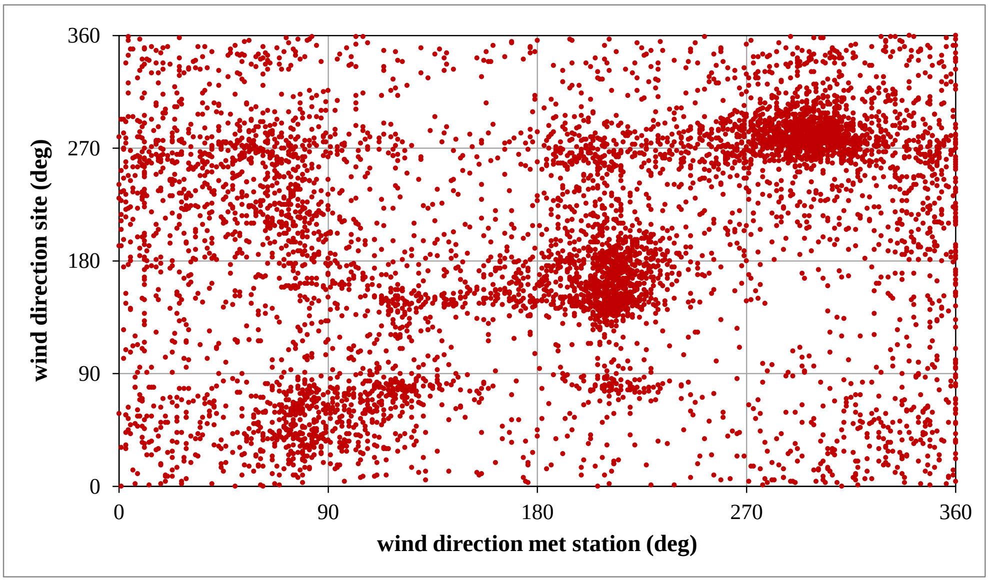

Figure 7: Concurrent hourly wind direction data corresponding to the wind speed data of Figure 6

Several clusters appear in the comparison of the wind direction data. The individual analysis of the clusters reveals that each of them is connected with a particular weather situation. Two of these are illustrated in the following.

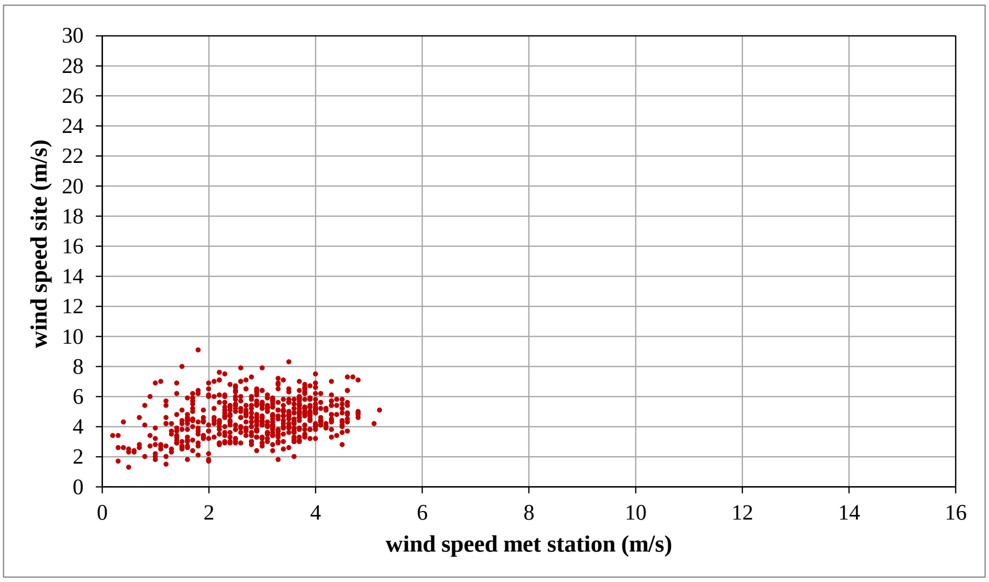

Figure 8: Data from Figure 6 only for winds from south-southwest at the met station and from south on site

The combination of winds from the south-southwest at the met station and the south on site only occurs during the afternoon, in all seasons but with a strong emphasis on the summer. This means that it only occurs during specific weather situations. At the site, the wind speeds are then around 4 m/s. At the met station, they are low with varying strength, but not correlated with those at the site. However, it is clear from Figure 8 that only a specific combination of wind speeds out of Figure 6 is linked with this situation.

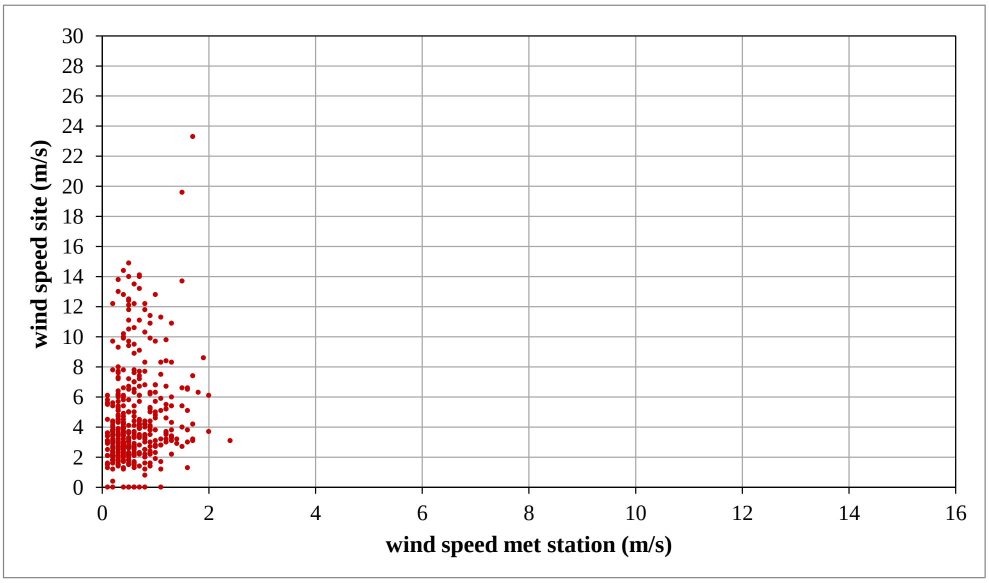

Figure 9: Data from Figure 6 only for winds from east-northeast at the met station and from west on site

It can be recognised from Figure 7 that the situation shown in Figure 9 only represents a relatively small proportion of time. However, it is not a random or erratic situation, like all others it is part of the range of situations that occur depending on the weather situation. In this case, it is almost exclusively limited to the night times, and with an emphasis on the winter season, followed by autumn and spring. Whilst regional temperature differences were the drivers for the winds at the site in this situation, making them pass above several chains of mountains from inland to the sea, these forces did not affect the wind flow at the met station due to its location in the valley. The winds simply passed above the valley. For the met station, temperature differences on the surrounding mountains (e.g. cold air masses flowing down mountain flanges) were presumably the main drivers of the wind during this situation. This created very low wind speeds at the met station, whilst the magnitude of the winds at the site depended on the temperature differences between the land masses and the Mediterranean Sea, but perhaps also on large scale patterns. The winds at the met station and the site were therefore independent from each other. However, they were not random, but limited to certain ranges of wind speed and direction.

Clearly, common MCP procedures are fully inadequate for such cases, which are then sometimes left without an acceptable means for the long term extrapolation. However, if a procedure was capable of providing a proper solution for such cases, this would not just be yet another (better or worse) variant of known methods, but it would really resolve fundamental problems for many projects.

3 Basic idea and concept of the MDP method

The MCP methods basically assume an idealised situation where for a given wind direction sector, a given wind speed of the reference data set corresponds to one well defined wind speed at the site.

The MDP method starts from the thought that a given combination of wind speed and wind direction at the reference station may well, for good reasons, correspond to a range of situations on site. It may therefore be preferable to maintain this range of situations, i.e. not to reduce it through averaging.

In the case of Figure 3, the wind directions of the reference and on-site data sets will usually be very similar, but a given wind speed at the reference station will correspond to a (limited) range of wind speeds depending on the atmospheric stratification. If the distribution of atmospheric stratifications during the year of the on-site observations is sufficiently similar to the long term distribution, the frequency distribution of wind speeds observed at the site for a given wind speed and wind direction at the reference station can be assumed to occur similarly in the long term. It will be sufficient to weight this distribution with the ratio of the frequencies of occurrence of this situation at the reference station between the long term and the period of concurrent data.

In the cases of Figures 6 to 9, a given combination of wind speed and direction at the met station will correspond to a quite variable range of wind situations at the site. However, if again it is assumed that the distribution of weather patterns that have occurred during the concurrent period between the site and the met station is representative for the long term (which at least requires that only periods of 12 months of concurrent data or multiples of these are considered), the entire range of situations observed at the site for a given wind situation at the met station may be weighted corresponding to the long term proportion of this wind situation at the met station.

This is what the MDP method does. For each combination of wind speed and wind direction in the reference data set, the matrix of wind speed and direction observed at the site is established. These matrices are then weighted according to the long term proportions of the corresponding combinations of wind speed and wind direction. Then they are added up, which yields the assumed long term frequency distribution of the winds at the examined site.

For Figure 3 for instance, the matrix of observed winds at 100 m height for a wind speed of 3 m/s (+/- 0.5 m/s) and a wind direction of 240° (+/- 15°) at the 10 m height anemometer is as follows:

	wind speed … (m/s)																				
wind direction (deg)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	0	5	11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
210	0	0	0	1	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	0	0	0	129	242	133	195	187	113	64	14	1	0	0	0	0	0	0	0	0	0
270	0	0	0	19	58	63	94	106	93	42	10	0	0	0	0	0	0	0	0	0	0
300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
330	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
360	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 10: Matrix of 100 m height data for a given wind situation at 10 m height in the case of Figure 3

If the 10 m height wind measurement corresponded to the met station in this case, the above matrix would be weighted according to the ratio of counts in the 3 m/s, 240° bin at 10 m height between the long term and the period of concurrent data with the 100 m height measurement. Clearly, in the present case the 10 m height wind measurement was not operating for longer than the 100 m wind measurement, but this would be the case if this had been a met station.

In common cases and in particular in more complex cases, the patterns in the matrices are of course more diverse.

4 Practical example

The following example demonstrates the power and the stability of the MDP procedure. The data measured on a site in a relatively flat region was extrapolated to the long term using the data from three met stations of the region in turn. The wind resources of these met stations, even their wind roses, differed strongly from that of the examined site. One of the met stations was in a fairly disturbed environment in a city, one on the borders of a city and one directly at a coast.

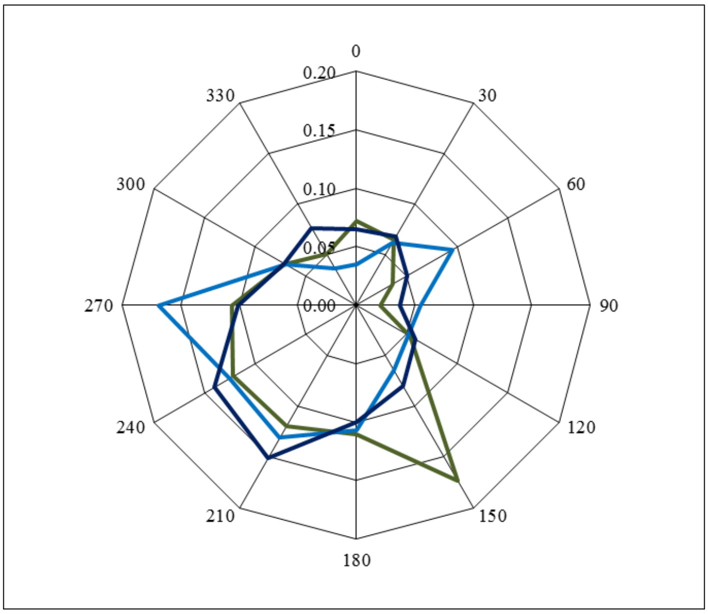

Figure 11: Long term wind roses of three different reference met stations

Since the long term wind field for the site is based on the wind field observed there, but accumulated from the frequency distributions for each weather situation with different weighting, the basic characteristics of the site are preserved with the MDP method.

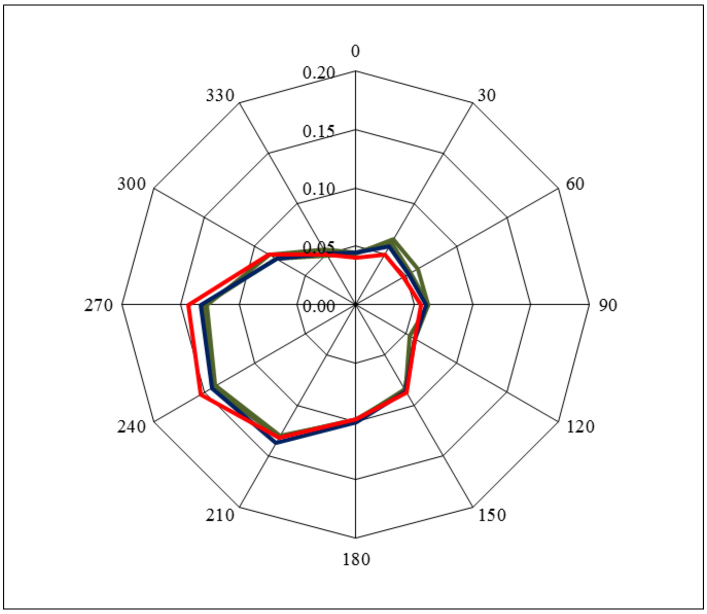

Figure 12: Wind rose measured on site (red colour) and extrapolated long term wind roses based on the data of the three reference met stations of Figure 11 (using the same colours as in Figure 11)

In the present example, the short term wind rose observed on site differs from all long term wind roses obtained via the MDP method, but these are very similar to each other. This means that the MDP method produced consistent results based on all reference data sets, although these are very different from each other and they differ from that of the site. On the other hand, the results for the site are not just a replication of the data of the short term period.

The following graph corresponds to Figure 12, but with respect to the frequency distribution of the wind speed.

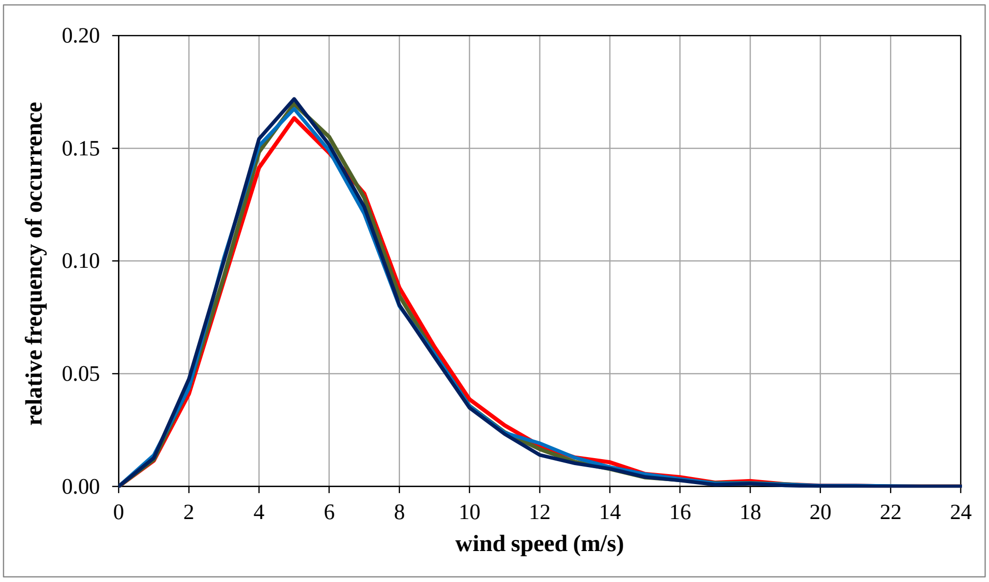

Figure 13: Frequency distribution of the wind speed measured on site (red colour) and extrapolated to the long term based on the data of the three reference met stations of Figure 11 (using the same colours as in Figure 11)

Although the difference between the short term and the long term is relatively small in this case, it is obvious from Figure 13 that, as with the wind rose, the differences between the long term frequency distributions obtained with the MDP procedure via a range of met stations are smaller than the difference between these and the observed short term distribution. The qualitative changes obtained through the extrapolation process are visibly similar. In quantitative terms, the three long term frequency distributions, applied to the power curve of a wind turbine, lead to a range of 4 % in long term energy production. Between two of the results, the difference is even just around 1 % (depending on the turbine type).

This example demonstrates that the MDP method is capable of producing consistent results which preserve the characteristics of the site.

5 Further notes

Since the MDP method is based on a statistical accumulation of matrices, the time stamps are lost. The method cannot be used to reconstruct time series.

Clearly the MDP method can in principle be applied to any arbitrary combination of reference data and on-site data. It is the responsibility of the user to determine upfront whether a strong meteorological connection exists between the wind at the site and the reference data set. Otherwise the result is meaningless.

Generally, the uncertainty of the method increases with the difference between the wind resources of the observation period (period of concurrent data) and the long term. The main reason is the assumption that the relationship between reference data and on-site data observed in the concurrent period corresponds to the long term relationship. This assumption is the same made by any MCP method as well. If the weather situation of the observation period differs from the long term weather situation, this may affect the relationship referred to above. Furthermore, with increasing number and strength of effects which cause a difference in wind resource between the reference and the examined site (i.e. stratification, local topographic and / or thermal influences, regional influences), the uncertainty of the method increases since these effects may have had a different magnitude during the observation period and the long term. Again, this also applies to the MCP methods.