

Solar Power Interpolation and Analysis using Spatial Autocorrelation

Syed Muhammad Ahsan Gillani

Chair of Geoinformatics, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Munich, Germany

Key words: Solar power interpolation, Solar power attenuation, Spatial autocorrelation, Semi-variograms, Geostatistics

Abstract

To reduce solar power production invariance, it is critical to study varying patterns of power production in the concerned region. Analyzing the patterns of past power production trends can help simulate power production scenarios for future. The current study area is around Amsterdam, located in Netherlands. PVoutput.org website is used to mine 6 months of solar power production data for 120 stations around Amsterdam city. FME Workbench software is used to actively fetch the data from the mentioned website and manage in a MySQL database. Solar attenuation maps created using ArcGIS, helped to graphically visualize the variations in solar power production at different times and locations. Further, spatial autocorrelation is checked between the stations using semi-variograms in geostatistical tool of ArcMap. This feature allows to check whether the stations located close to each other are more correlated to each other rather than stations which are far apart. The statistical data analysis of power production can aid solar power production companies to better interpolate and predict solar power in advance for the concerned study region.

Corresponding Author: Syed Muhammad Ahsan Gillani  Muhammad.gillani@tum.de

Foreword

The stable power sources; coal, gas and nuclear plants, have long ruled the power distribution process which are relatively simple to handle by grid operators. The fluctuating renewable energy (Photovoltaic, wind power etc.) input into grid, on the other hand, puts an extra pressure on grid operator to manage it with stable power sources. Increasing demand of energy motivates for integration of solar power into existing energy systems. Solar Power is acknowledged as inexhaustible and most favorable renewable resource for bulk power production. Photovoltaic panels transform solar energy into electric energy (Wan, et al., 2015).

Solar power production data analysis and forecasting is an important process to efficiently manage electricity grid plus to perform solar energy trading. This requires basic understanding of Sun path, atmospheric conditions, scattering phenomena and most importantly the analysis of power data produced from photovoltaic plants (Singh, 2013). The main test for electrical grid operators is to harmonize, continuously, the demand with energy supply (R.A. Verzijlbergha, 2017). For valid and productive supply of modern power systems; real time, 1 day ahead, a week, month or year earlier schedules are being made. It enables to foresee power system load curves and outputs of PV power plants (Kou, 2014). The Power Production/Performance data are provided by many photovoltaic system users on websites including pvoutput.org. The website pvoutput.org allows users a platform to input actual-time data from their PV systems available to the public and to compare it with other systems. The time resolution of the power data is up to 15 minutes and is updated accordingly. Because of several active users who report the power data online, it has eased the evaluation and prediction of solar power data in the concerned region.

Motivation of Study

The integration of solar energy and precise prediction can aide in reliable management and distribution of energy. In this regard an immense analysis of past solar power production data and its correlation with weather parameters is needed. (Chaturvedi, 2016). The purpose of this study is to provide solar power producers with optimal power production location and times for electricity generation from Photovoltaic systems.

The focus of the project is therefore to collect performance data from several stations around Amsterdam over a longer period, to refine it to get meaningful comparison. This performance data of a station can also be visualized in terms of space, along with various other stations to observe the phenomena of spatial auto-correlation. It can be checked that whether PV stations located near are more correlated in terms of solar power production, than the stations that are far from each other. This spatial autocorrelation can be checked using semi-variograms, a geostatistical feature in ArcGIS. Prediction of solar power; considering the current power production data and weather forecast is a great inquisition now-a-days for electric utilities, which can be attributed to future works.

Data Collection and Analysis

The study site is area around Amsterdam city, from where the solar panel stations are recorded using PV output.org website. Each station along with its name, co-ordinates (latitudes/longitudes) and a specific station id; is used in FME Desktop to fetch power output data from the website. FME further eased the process, with no need of extensive coding, to transform data in MySQL database (Safe, 2018).

SQL; Structured Query Language (Microsoft, 2017) is a programming language used in MySQL Workbench to organize data held in a relational database management system (RDBMS). In MySQL, a table is created in which all station names, station ids and their location are recorded.

FME Workbench interface is used for parsing HTML of the PV output website and transferring solar power production data to MySQL database. Out of the various data available for a station; Energy, Efficiency, Temperature, Power, etc., the study was concerned with Date, Time, and Power. Thus, FME was instructed to only retrieve the relevant data that is to be further analyzed for power correlations. At the very end of the FME flow diagram, formulation was done to transform the extracted parameters i.e. power production to MySQL Databases. The process is continued until a 6 months solar power production database was maintained.

Accessing frequent data through parsing, blocks access to pvoutput.org after each 12 days of data retrieval for each station. For this purpose, several Gmail accounts are used for continuous retrieval of data. The overall process takes time to fetch database of several months. After every 12 day of data extraction, a refresh button was clicked in MySQL to update the latest solar power data. After the successful data acquisition, attenuation factor is calculated which is used as a basis to perform solar power interpolation.

Defining Attenuation Factor

The multiple factors which influence amount of solar energy produced (Elfouly, Donaubaue, & Kolbe, 2018), are listed as follow

Solar Orientation

Time of the year, time of day (Hoste, Dvorak, & Jacobson, 2009)

Location, topography

Alignment and tilt of the system

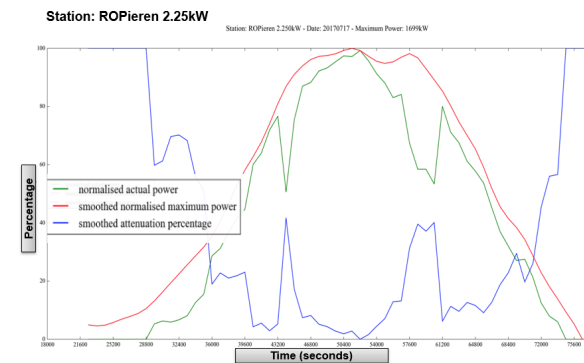
Weather (Cloud-cover, Temperature, Wind, Air pressure...)

Maintenance (Air pollution i.e. airborne particles)

The main idea is to evaluate the large sum of Photovoltaic power measurements statistically, rather than analytically. Because analytical considerations of all above mentioned influences, involving complex formulas and calculations, is practically impossible (Elfouly, Donaubaue, & Kolbe, 2018). Sun position is one major influencing factor in solar power production and changes frequently in a day, hence, a high-resolution data becomes a need to have a reasonable forecast of solar power. In winter, when the sky is clear, of course, due to the lower position of the sun, significantly less energy is generated than on a sunny summer day. A correlation with the degree of cloudiness is therefore not meaningful in this form. There is a need of energy data normalization first in a way that they are comparable over all day and seasons (Markus, 2017). For this purpose, the so-called Attenuation factor (damping factor) is introduced.

Attenuation factor defines the weakening of the ideal solar power production in percent. The attenuation is the factor by which the actual power is lower than the maximum solar power (that can be generated under ideal conditions). The ideal conditions are ought to be clear sky, least airborne particles present in air and ideal solar

orientation which results in maximum solar radiation transfer to the surface (Mani & Chacko, 1980). The ideal power let us calculate a reference power curve indicating maximum power that could be produced in a day. In summer, the reference line is naturally higher than in winter and at midday larger than in the evening. The attenuation factor is independent of the time of the year as it relates the actual and the reference power hence a normalized quantity (Markus & Lukas, 2017).



The above figure shows the variation of power production and attenuation factor throughout the day time on July 11, 2017. Further, it needs to be elaborated how the attenuation factor is calculated. The Red Line is the reference line; it sets the maximum producible energy output to one clear day of the sun, which is found by looking power values in a window of 30 days around the date 2017-07-11 (Elfouly, Donaubaue, & Kolbe, 2018). The green line shows the actual generated energy output in the station. Attenuation (blue line) then results from the formula.

$$A(\%) = \left(1 - \frac{P}{P_{max}}\right) * 100\%$$

In the formula, P is the measured power at a distinct point in time, while Pmax is the maximum possible exploitable power at this point in time during clear sky conditions. Again, this maximum power is calculated from the statistics. The assumption made is that there exists at least one full sunny day within a period/window of 30 days (Elfouly, Donaubaue, & Kolbe, 2018).

Solar Interpolation in ArcGIS

The interpolation method chosen for this study is Natural Neighbor Interpolation (also called Voronoi Interpolation). This method has advantages over simpler methods of interpolation, such as nearest-neighbor interpolation, in that it provides a smoother/continuous interpolation surface and hence more appropriate to define spatial variations of attenuation (Elfouly, Donaubaue, & Kolbe, 2018).

First, a base map of Amsterdam was loaded on ArcGIS, so that any layer loaded over the base map can be seen according to its position on earth. The stations data table in MySQL, containing their coordinates databases, is transferred to shapefile to locate the stations on the base map. FME Workbench to transform solar power production data to shapefiles using shapefile creator feature.

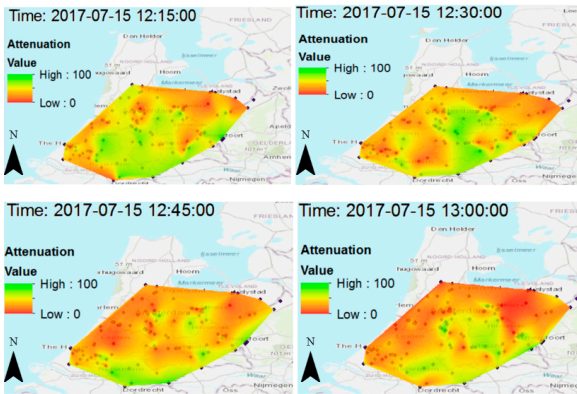


For performing solar interpolation, shape files are produced for each point in time and are created using FME Shape file extension. This loads the interpolation data from MySQL databases and transfers it into shapefiles ready to be used by ArcMap. ArcMap enables the user to extract data from shape files and plot various desired parameters (attenuation) of a geometrical object (stations). The spatial coordinate system used while transformation is the common LL84 system. FME Shapefile writer has a very interesting feature that enables one to produce shapefile name according to any of the feature value, here time. For example, the shape file name 201706240600 refers to June 24, 2017 and 06 AM.

Each of the shape file containing attenuation values of all stations, for a particular time, is loaded into ArcMap interface. To interpolate the attenuation in between the stations, where no values of attenuation are present, natural neighborhood attenuation is employed. The interpolated data can be seen in the form of varying colors to differentiate between attenuation values. It is to be noted that attenuation values of all stations referring to a particular time, have different geometrical locations on the map. For interpolating the attenuation values in the area between these stations, natural neighborhood interpolation is selected from Arc-tool box and all the shape file layers of a day are loaded into the dataset. Interpolating using the loaded layers; result in colored interface of attenuation values varying in the whole map containing the 120 stations.

It proves to be a very helpful tool as it helps to analyze which location is best to install solar panels. And, one can know that at which time of the day maximum power can be harnessed.

A very organized tool of ArcMap is ‘Model builder’ which helps to collectively perform several individual steps and hence reducing the time of processing. It consists of workflows that combines array of geoprocessing tools to create, edit and manage models. Model Builder can also be considered as a visual programming language for building workflows supplying output of one tool as input for the next in line (What is Model Builder?, 2014). This tool was used to produce a collection of interpolated images for all time of the day. After the creation of mosaic image dataset of above mentioned interpolated images, time field is added to the attribute table of image file, to produce a video of all attenuation images combined using time slider tool.



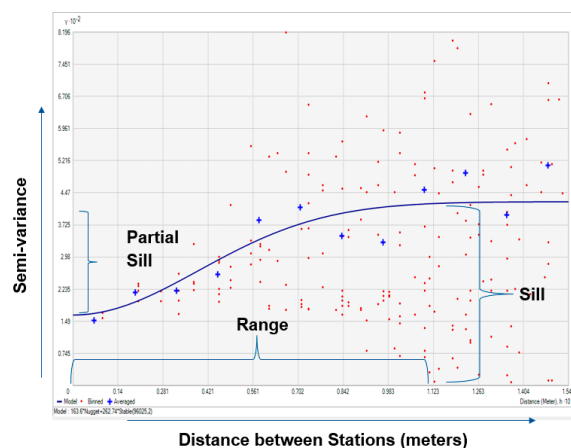
In the figure above, attenuation variation of a midday time on July 15, 2018 is shown. The areas surrounding PV stations are marked with color ranging from red (lowest attenuation) to green (highest attenuation). Attenuation seems to be decreasing at every next time interval starting from the time 1215 to 1300. This shows that with time more solar radiation is reaching the panels. Hence indicating a clear sky, which will result in higher power production at PV panels at 13:00 as compared to the time 12:15.

Now feed-in-power for any station can be predicted at a specific point in time and space using attenuation field as reference indicator. This use case is related to a real-world problem, that energy companies face. Only few stations could afford the installation of smart meters, which production companies require to predict feed-in-power for all stations connected to the network.

Hence, to predict the potential of solar power production for stations without smart meters, attenuation field can be a helpful parameter for extracting attenuation values at different geographical locations in the field. For solar power forecast, again, all these results heavily depend on the unpredictable weather. A high frequency weather database (i.e. cloud cover) can aide to forecast attenuation for near future.

Spatial Autocorrelation Using Semi-Variograms

The semi-variogram displays the spatial autocorrelation of the measured sample points (here stations). The variogram is a discrete function calculated using a measure of variability between pairs of points at various distances. As each pair of locations is plotted in the variogram, a model is fit through them (Understanding a Semivariogram, 2014).



Each red dot indicates two stations in fact. X-axis shows the distance between two neighboring stations. Higher is the distance between stations that are located far away, shown here on the far right of plot. Noticing variogram, the point on x-axis (distance) where the model first straightens out is called the range. Sample locations separated by distances less than the range shows a spatial autocorrelation, but locations further than the range are not correlated. The value of semivariogram on y-axis at the point of range is called the sill. Nugget is defined as the point where fit curves cuts the y-axis while the partial sill is obtained by subtracting nugget from sill (Understanding a semivariogram: The range, sill, and nugget, 2014).

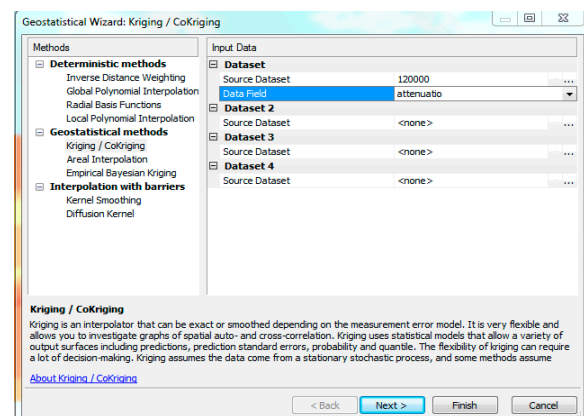
A semivariogram value (squared difference of attenuation between the values of each pair of locations) is plotted on the y-axis while x-axis plots distance separating each pair of measurements in meters. Multiply the x-values by 10E4 to get the exact value of difference between stations.

The technique considers the semi-variogram values between two location points as follow:

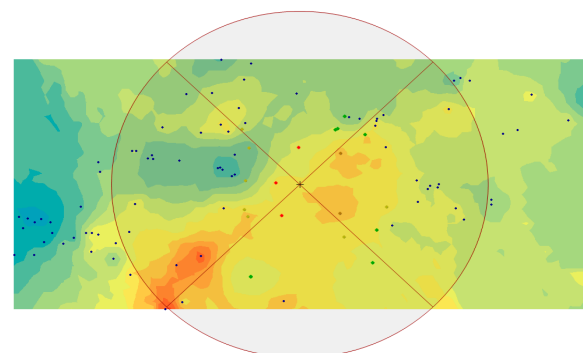
$$\text{Semivar.} = 0.5 * \text{average}[(\text{value at location } i - \text{value at location } j)]^2$$

Using this technique, it is checked whether the attenuation (at a certain time of a day), among neighboring stations, is correlated with distance or not. In this regard a particular station data at a certain time was converted into a shapefile using FME Workbench. This shape file was loaded into ArcMap interface as was done in the case of doing interpolation. The shapefile loaded is of June 24, 2017 12PM, which contains the features attenuation, latitude, longitude and date/time fields.

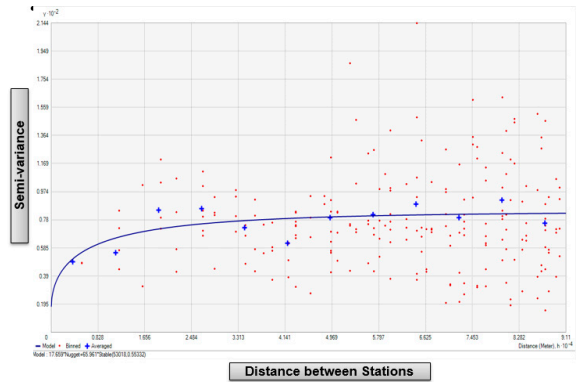
After enabling geostatistical analyst extension and spatial analyst extension, geostatistical wizard is run by clicking on the geostatistical extension. The following window pops-up:



In the above interface, attenuation parameter is selected to check for spatial autocorrelation among the station. In the next window (Figure 23) select Kriging Type as Ordinary and output surface type as prediction. The Kriging/Co-Kriging method check spatial correlation between the neighboring points on map by dividing the target area into sectors of a circle.

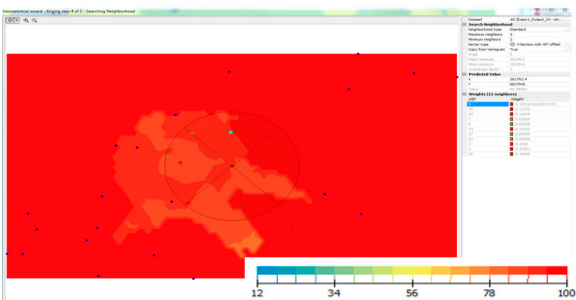
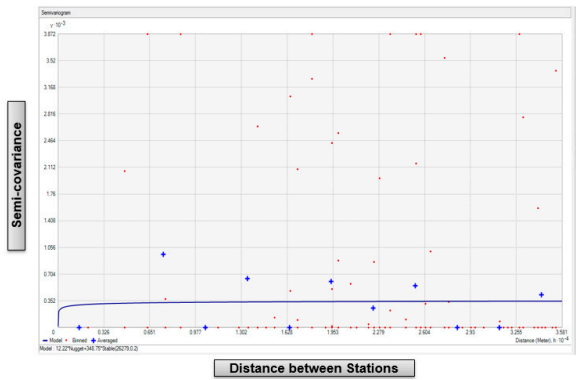


The following are the variogram results for June 24, 2017, 12 PM:



From the above plot, it is evident that solar power attenuation variates in the region. High values of semi-variance values indicate that the solar power production variates in the region. However, stations located close together are more spatially correlated than the ones far away.

During the wee hours (6 AM to 8 AM), mostly there is no variance in attenuation seen. This is because of no solar radiation reaches earth, hence attenuation is maximum and same throughout the region. For an example, June 24, 2017 06 AM time is chosen, and the following plot is obtained after performing geostatistical analysis.



During early hours, there is no solar energy reaching the earth and hence highest attenuation.

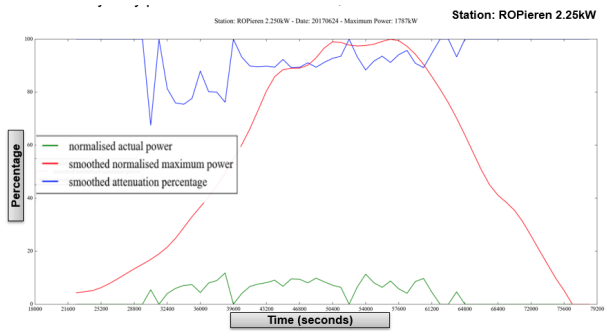
The attenuation remains same throughout the region, hence, minimal and same semi-variance throughout the area. There exists spatial autocorrelation i.e. the attenuation at all the stations is same in space but least power production.

Results and Discussion

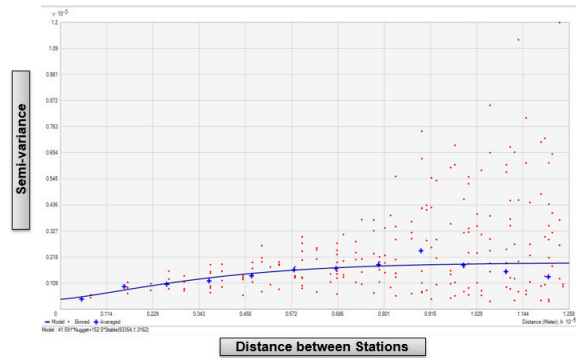
Further analysis was done between a cloudy, clear sky and moving cloud day to see how PV power stations correspond to solar energy in reference to space.

Cloudy Day

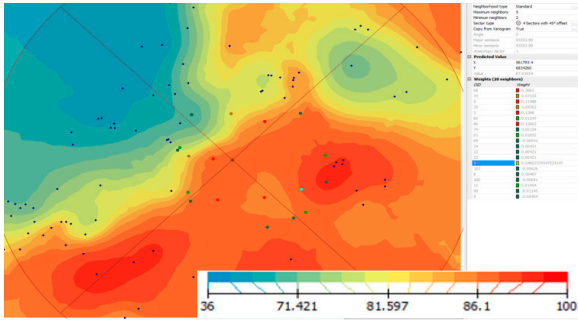
If the actual power is very less as compared to the maximum power produced, at a certain point in time, one can ascribe that time (day) to be covered with clouds. The following plot, for station name ROPieren 2.250kW, confirms the stated argument, showing a cloudy day on June 24, 2017.



Power data from June 24, 2017 is also loaded into Arc GIS as shape file and geostatistical analysis is performed on the data, at 1500 (54000 seconds), which is shown in the following semivariogram:

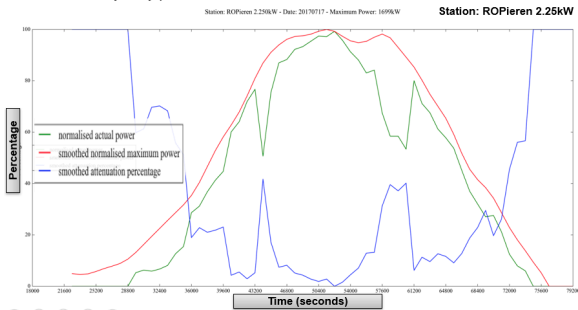


It can be seen that due to cloud cover, semi-covariance values are less and doesn't change much. A high attenuation (mostly over 70%) can be interpolated, which is also confirmed by the following neighboring plot of geostatistical wizard as follow:

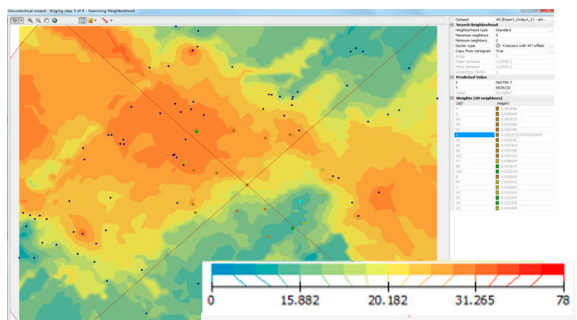
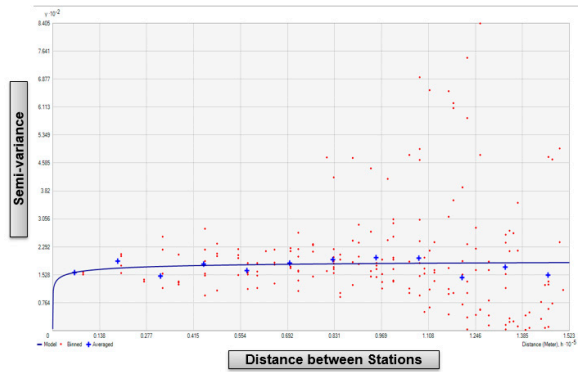


Clear-Sky Day

Now a clear sky day is plotted for the date July 17, 2017. Clear sky condition on this day is evident by looking at the day curve of the station 1 (Station ROPieren 2.250kW). Actual power produced matches closely with maximum power for most time of the day.

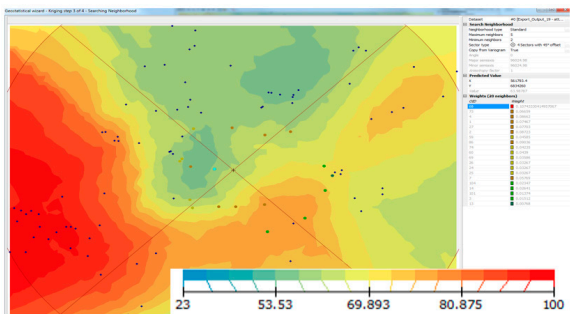
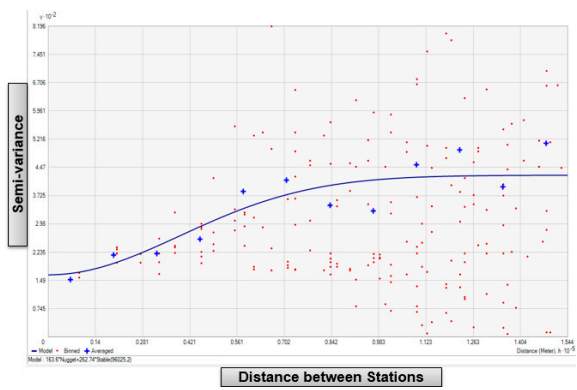
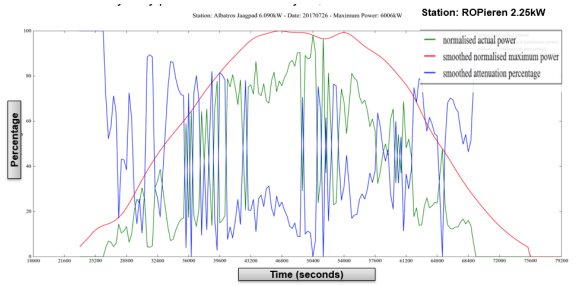


It is further found out that during July 17, 2017, the time 1445 shows clear sky time, therefore minimum attenuation at station 1 and surrounding area. The attenuation values are less than 30% and does not differ much because of clear sky conditions, observed in the whole area. Minimum semi-variance values are observed among stations. This is also evident from the following plots:



A Day with moving clouds

Several data-plots were searched in the power database created, out of which a day with moving clouds, July 26, 2017, is found in Amsterdam. The day represents high power variation which led us to believe that it's a moving cloud day. Following productive results are found:



Quite a variation in attenuation values ($23 < \text{attenuation} < 100$) is seen throughout the map, resulting in increasing semi-covariance values. The stations located close together are more spatially correlated to each other than the stations located far away. In case of moving cloud day, the solar power production will be highly unpredictable in the concerned region, hence power utilities must be very cautious to manage this uncertain power with traditional power supply networks.

Conclusion

Provided a large solar power database, solar power interpolation and spatial auto-correlation analyses can be done in any concerned area. Using this study, any other solar power production data can be organized using smart approaches to manage data through MySQL and FME Workbench. Further analysis of data and inclusion of data as shape files into ArcGIS can allow the user to efficiently play the data in terms of colorful visuals, which enables better interpretation of solar power variation. All seasonal variations can be accounted if attenuation factor is used, as it is independent of the time of the year. It is found by the difference between reference power (maximum power), calculated using a 30-day moving window concept, and actual power. Solar Power interpolation maps made at various locations and different time intervals can aid solar power production companies to locate new PV panels where maximum power can be harnessed. If provided with meteorological data, power utilities can also predict future power production and hence can ensure sustainable grid management.

Spatial autocorrelation using semi-variogram in geostatistical analysis tool in ArcGIS, allows to check for spatial description and spatial prediction. Solar Power utilities can already predict to some extent how power production will vary in case of a moving cloud day, sunny day or cloudy day. In case of cloudy, clear sky and early hours' time, low semi-covariance values are observed because of similar values of attenuation in the region. On the contrary, on moving cloud day high semi-covariance values are observed due to high variation of attenuation values between the PV stations in the region. Additionally, on a moving cloud day, stations located close together are found more correlated than the stations located far away.

However, on a moving cloud day, power utilities must be very cautious to manage the fluctuating power production in the overall concerned area. With above mentioned solar interpolation and spatial autocorrelation tools applied to any further region, one can produce a comprehensive understanding of time and location based solar power variation.

References

- Chaturvedi, D. K. (2016).** Solar Power Forecasting: A Review. *International Journal on Computer Applications*.
- Elfouly, M., Donaubauer, A., & Kolbe, T. H. (2018).** Prediction of feed-in power from photovoltaics in local distribution networks. *Wissenschaftlich-Technische Jahrestagung der DGPF und PFGK18 Tagung. München: Deutsche Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*
- Hoste, G. R., Dvorak, M. J., & Jacobson, M. Z. (2009).** Matching Hourly and Peak Demand by Combining Different Renewable Energy Sources, A case study for California in 2020. *Stanford University Department of Civil and Environmental Engineering Atmosphere/Energy Program*.
- Markus, E., & Lukas, R. (2017).** Nowcasting of Attenuation Values for Photovoltaics. *Final Report of course Geoinformatics II, Technical University of Munich*.
- Microsoft. (2017).** Retrieved from <https://docs.microsoft.com/en-us/sql/odbc/reference/structured-query-language-sql?view=sql-server-2017>
- R.A. Verzijlbergha, L. D. (2017).** Institutional challenges caused by the integration of renewable energysources in the European electricity sector. *Renewable and Sustainable Energy Reviews*, 660-667.
- Safe. (2018).** Retrieved from <https://www.safe.com/fme/fme-desktop/>

Singh, G. (2013). Solar power generation by PV (photovoltaic) technology: A review. Energy, 1-13.

Understanding a Semivariogram. (2014).

Retrieved from ArcGIS:

<http://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/understanding-a-semivariogram-the-range-sill-and-nugget.htm>

Understanding a semivariogram: The range, sill, and nugget. (2014). Retrieved from ArcGIS: <http://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/understanding-a-semivariogram-the-range-sill-and-nugget.htm>

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., & Hu, Z. (2015). Photovoltaic and solar power forecasting for smart grid energy management. CSEE Journal of Power and Energy Systems, 38-46.

What is Model Builder? (2014). Retrieved from ArcGIS: <http://desktop.arcgis.com/en/arcmap/10.3/analyze/modelbuilder/what-is-modelbuilder.htm>