Master Thesis Proposal

**How to Distribute New Solar Systems in Europe to Reduce Power Generation Variability**

Department Environmental Systems Science / Earth Sciences, ETH Zürich

Institute for Atmospheric and Climate Science

Supervisors:

Dr. Jan Wohland, Institute for Environmental Decisions, ETH Zürich

Prof. Dr. Martin Wild, Institute for Atmospheric and Climate Science, ETH Zürich

Dr. Doris Sylvia Folini, Institute for Atmospheric and Climate Science, ETH Zürich

Dr. Stefan Pfenninger, Institute for Environmental Decisions, ETH Zürich

Submitted by

Dirk Mühlemann

13-732-037

Uster, 31/05/2020

Summary

To reduce greenhouse gas emissions and combat climate change, the electrical power production sector is facing a fundamental transition from conventional fossil to renewable technologies. The transition has already started, which can be seen by the great effort and ambitious targets of many nations around the globe. Installed power production capacities of solar photovoltaics (PV) are increasing every year and are already capable of producing around 5.5% of the Europeans electricity demand. Since PV power output depends on weather and climate, it exhibits a highly variable production pattern. This variability challenges the electricity grids because the stability of the grids depends on balanced supply and demand. Further massive deployment of PV systems could lead to an increase of the variability and therefore add to this challenge. Different studies suggest strategies to reduce the PV power output variability on rather short (minutes to hours) or long (seasonal) timescales. However, weather regimes lasting several days influence the PV power output across Europe as well, and knowledge of its impact on multiday PV power output variability is still limited. In this thesis, we aim to identify spatial distributions of newly installed PV systems that minimize the multiday power output variability within Europe. To quantify the variability, we will perform empirical orthogonal function (EOF) analyses of solar downward radiation, which influences the PV power output directly, and geopotential height at 500hPa which reflects weather regimes and therefore has an indirect effect on the PV power output. Both fields will be taken from the ERA5 reanalysis dataset which covers the time period from 1979 to present. The resulting subspace spanned by the leading EOFs of our analysis will be grouped in different weather regimes with the k-mean clustering techniques. To assess the PV power output variability the resulting regimes will be related to country/region specific PV capacity factors, which are defined as the quotient of actual power output and installed PV capacities. We will use hourly PV capacity factors provided by the simulation of renewable.ninja from 1985-2019. Connecting the regimes with the PV capacity factors will lead to an overview of under- and overproduction (relative to the mean) per country/region and weather regime. The current installed capacity of PV systems in Europe will then be used together with our findings to assess the current multiday PV power output variability in Europe. Furthermore, an optimal distribution of additional PV systems will be proposed with the goal to minimize the multiday variability.

Contents

[1. Background 4](#_Toc41830551)

[2. Objectives 6](#_Toc41830552)

[3. Data & Methods 7](#_Toc41830553)

[3.1 Data 7](#_Toc41830554)

[3.1.1 ERA5 7](#_Toc41830555)

[3.1.2 Renewables.ninja and the Global Solar Energy Estimator (GSEE) 7](#_Toc41830556)

[3.1.3 Installed PV capacities 10](#_Toc41830557)

[3.2 Method 10](#_Toc41830558)

[3.2.1 Weather regime classification 10](#_Toc41830559)

[3.2.2 Capacity factor estimates 11](#_Toc41830560)

[4. Timeline and milestones 13](#_Toc41830561)

[5. References 14](#_Toc41830562)

# Background

Many governments undertake ambitious climate mitigation efforts to reduce the adverse effect of global warming and thereby try to meet the 1.5°C goal from the Paris agreement (Hulme, 2016). To achieve this goal a transition from conventional fossil to renewable energy technologies is substantial. Solar power generating photovoltaic (PV) systems, as one of the major renewable technologies, has seen a tremendous growth in recent years. According to the PV status report (2019) by the European commission, there was a global installed capacity of 520GW by the end of 2018. By the end of 2019 the installed capacity is expected to reach 650GW, which allows the PV systems to produce roughly 4% of the global electricity demand. Europe's share of installed PV capacity at the end of 2018 was about 117GW, which allowed to produce 5.5% of Europeans electricity demand. Furthermore, recent scenarios for the necessary installed PV capacity if the world is to reach 100 % renewable electricity production in 2050, suggest that the PV installed capacity must rise to more than 4TW by 2025 and 21.9TW by 2050. For Europe this would imply an PV installed capacity of 630GW by 2025 and 1.94TW by 2050 (Jäger-Waldau, 2019).

PV power output depends on weather and climate and therefore challenges the current power grid by variable power input (Graabak and Korpås, 2016; Stram, 2016). Within minutes, the power production from a PV system can vary by as much as 80% due to passing broken cloud cover (Mills and Wiser, 2010). There is also a diurnal cycle with highest power production during the day and no production during night. By analysing multiday fluctuation, a relation between long lasting weather regimes and PV power production can be observed. During one weather regime the PV production pattern varies substantially to the next weather regime (Grams *et al.*, 2017). Furthermore, variability in the PV power production is caused by the seasonal cycle (Heide *et al.*, 2010). The efficiency of a PV panel is mainly dominated by the available surface solar radiation and the panel temperature (Huld *et al.*, 2010). Since the seasonal cycle affects both, a direct effect can be observed. Even decadal to multidecadal analyses suggest that global warming and solar brightening/dimming influences the PV power output. Solar brightening/dimming is often caused by air pollution patterns, which influences the available surface solar radiation and therefore affects the PV power output. Global warming leads to an increase of panel temperatures and reduces the efficiency of the panels (Wild *et al.*, 2015; Sweerts *et al.*, 2019).

A stable power grid depends on balanced electrical supply and demand (Stram, 2016). Short time scale variability due to passing broken cloud cover can affect one PV system substantially. But wider geographical distribution of multiple interconnected PV systems can significantly smooth this short time PV power output variability (Delucchi and Jacobson, 2011; Graabak and Korpås, 2016). Co-deployment of renewable energy system (water, wind and solar) can counteract the variable power output caused by diurnal and seasonal cycles. The basis for these approaches is the different diurnal and seasonal production pattern of water, wind and solar power plants. For example, wind power output exhibits highest production rates during winter whereas solar power production is highest in summer. Co-deployment of wind and solar can therefore reduce the seasonal power production variability. Other combination are possible as well (Heide *et al.*, 2010; Santos-Alamillos *et al.*, 2015; Graabak and Korpås, 2016). To summarize, different studies already have proposed methods to reduce short, diurnal or seasonal solar power production variability.

Fewer studies have investigated in reduction of multiday solar power output variability from a meteorological standpoint. Grams *et al.* (2017) did but rejected the idea to further invest in it, based on their findings that it would need a tenfold increase of installed PV capacity in Europe to be comparable to the variability of wind power output. Therefore, they focused their study on wind and concluded, that spatial deployment of wind fleets based on information of different weather regimes can reduce the wind power output variability within Europe substantially. Nevertheless, they showed that connecting weather regime to PV power output variability is also possible. Even though the decision to focus on wind rather than solar power output variability is comprehendible, calculations of necessary future installed PV capacities gives reason to do the investigations anyway. Ram *et al.* (2017) estimated that the installed PV capacity for a 100% renewable scenario in Europe must rise to 1.94TW by 2050. This is roughly a twentyfold increase of installed PV capacity compared to the 87.19GW installed PV capacity used in Grams *et al.* (2017) study. Therefore, the impact of multiday PV power output variability caused by different weather regimes could also become substantial, which makes investigation of the optimal spatial deployment of future PV systems in Europe before further massive deployment of great interest. The results could support current planning activities and could help to reduce future grid balancing problems.

# Data & Methods

## Data

### ERA5

For the weather regime definition, we use the reanalyse dataset, ERA5, which is published by the European Centre for Medium-Range Weather Forecasts (ECMWF). It provides atmospheric, land and oceanic variables from 1979 to present with a temporal resolution of an hour. The grid of ERA5 has a horizontal spatial resolution of 0.28 degrees (~31km) and 37 pressure levels in the vertical. A detailed documentation about ERA5 reanalyse dataset can be found on the ECMWF confluence web page (Hennermann and Yang, 2018).

We use the 500hPa geopotential height and surface downward solar radiation variables from ERA5 dataset in the domain 80°W to 40°E, 30°N to 90°N. The domain is chosen according to Grams *et al.* (2017) for easy comparison, and is also reasonable for our meteorological field investigates since it captures the large scale circulation that affects Europe. The following chapter describes the pre-processing of the raw ERA5 data which is needed before the weather regime definition can take place.

#### ERA5 data pre-processing

The temporal coverage of the initial dataset is on an hourly basis. Since we investigate in weather regime which last several days the dataset is reduced by calculating daily means of the geopotential heights at each grid point. The input data for our EOF analysis are standardized anomalies (EQ1) of these daily means.

where z\_all are the daily means of the geopotential height, climatology\_mean is the climatological mean for every grid point with a running window of XY days, and climatology\_std is the standard deviation with a XY running window.

The standardized anomalies are calculated with a XY running window for the reference climatology and a XY running window for the standard deviation. The running window is defined such the respective day acts as centre of the window. For instance, to calculate the reference climatology a separate mean for every day is calculated which includes XY days before and after that day. This results in 40 means for every day since our dataset covers 40 years. These 40 means are taken to calculate a mean again, so that we finally have 366 means for every day of the year as reference climatology. This is done analogous with the standard deviations. The standardized anomaly is used because we want to define weather regimes year around. Since the standardized anomaly includes normalization with the standard deviation, the amplitude in the anomaly caused by the seasonal cycle is removed (GRAMS) prior to the weather regime classification.

Nan 🡪 bfill and ffill????

Std 🡪 spatial mean???

## Method

### Weather regime classification

The EOF analysis on the standardized anomalies is performed with the eofs python package by Dawson (2016). The resulting first 15 principle components which explained ~90% of the variance are used to cluster the data into weather regimes. We use the clustering method k-means which is implemented in the python package sklearn.cluster by Pedregosa *et al.* (2011). According to GRAMS the optimal number of clusters to define weather regime year around is seven which seems to be plausible by a simple check with the elbow method:

EVTL. BILD

We will use the 500hPa geopotential height and surface downward solar radiation variables from ERA5 dataset. Before the EOF analyses and k-mean clustering will be performed, there is a need of preparing the data. The major reason to do so is that we want to understand the multiday variability and not diurnal or seasonal variability. Therefore, we will follow a similar approach like Grams *et al.* (2017) and adjust it were needed. This includes:

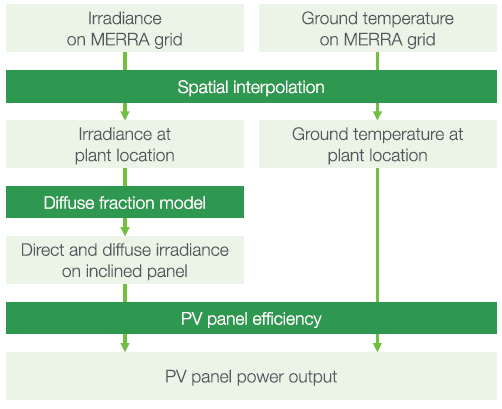
1. Separating seasons and calculating anomalies: Instead of absolute numbers we will calculate anomalies by using a running mean as reference climatology. We will begin with a simple 90 day running mean, similar to Grams *et al.* (2017) and potentially refine the filtering afterwards.
2. To ensure that the daily cycle does not dominate our EOF analyses, we need to filter the data beforehand. Grams *et al.* (2017) used a 10-day low pass filter and we will take this again as basis but will finally define our filter and its criterion during work with the data.

### Renewables.ninja and the Global Solar Energy Estimator (GSEE)

Renewables.ninja ([www.renewables.ninja](http://www.renewables.ninja/)) is an interactive web platform that simulates hourly power output of wind and solar power plants all over the world. To calculate the PV power output, it uses the so-called Global Solar Energy Estimator (GSEE). The source code of the GSEE is freely available on [GitHub](https://github.com/renewables-ninja/gsee) and a detailed description of the GSEE can be found in Pfenninger and Staffell (2016). The theoretical background of the GSEE is based on Huld *et al.* (2010). The following variables are the key input parameter of this method to estimate the energy yield of PV modules:

* Direct and diffuse irradiance at the PV panel
* Panel temperature

Pfenninger and Staffell (2016) used two data sources to estimate these variables: First the reanalysis dataset Modern-Era Retrospective analysis for Research and Applications (MERRA and MERRA-2) and second the Surface Solar Radiation Data Set - Heliosat (SARAH). Figure 1 shows the general approach of GSEE.



**Figure 1:** Overview of the approach used to model PV power output (Pfenninger and Staffell, 2016).

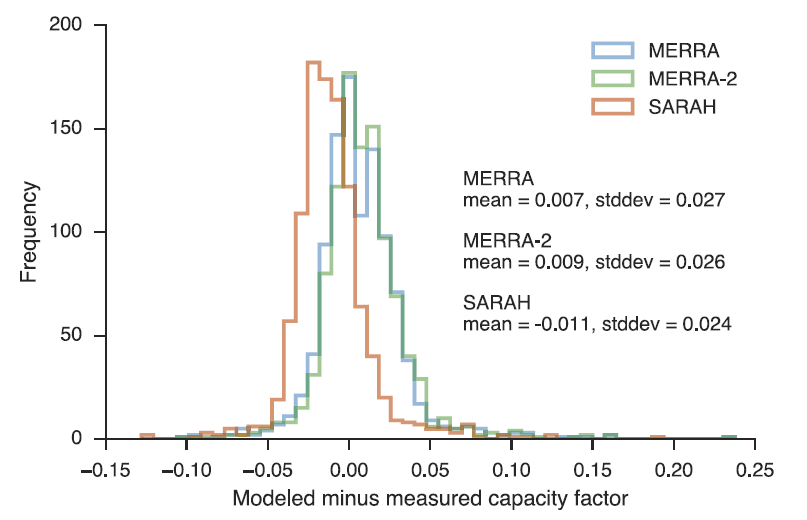
Both data sources provide the variables in hourly intervals. The spatial resolution of MERRA-2 is 0.5° latitude and 0.625° longitude. SARAH has a higher spatial resolution of 0.05° × 0.05°. MERRA only provides direct irradiance, but diffuse irradiance is needed as well. Therefore, they used the Boland-Ridley-Lauret model to estimate the diffuse irradiance (Ridley *et al.* 2010; Lauret *et al.* 2013). Since SARAH provides direct and global irradiance no further estimates were needed. Additionally, they use T2M (temperature at 2m above ground level) from MERRA-2 as estimates for the ambient temperature. To get the panel temperature they used the ambient temperature of MERRA-2 and additionally considered the effect of the irradiance on the panel temperature. This relation was estimated with site measurements of one of their sources (DTI see below). This dataset provides ambient and panel temperature for each site with which they derived an empirical relationship.

It is more common and easier to make comparison and analyses of PV power output with capacity factors rather than with absolute power output and we will use this approach as well. The unit-less capacity factor CF is defined as:

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 1 |

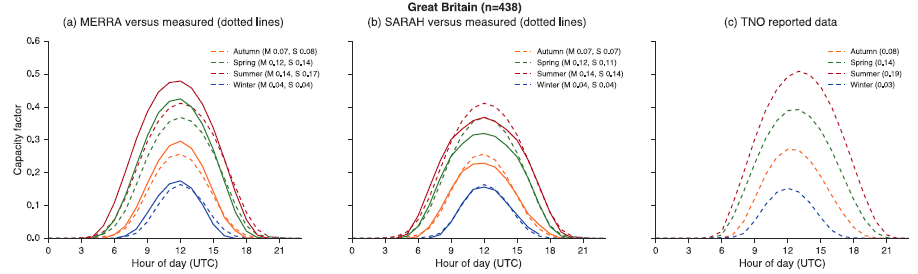
where P is power output [W] and IC is the installed capacity [W].

To evaluate the results obtained with the method described above, Pfenninger and Staffell (2016) compared it with capacity factors based on site measurements. To obtain the measured capacity factors they used three sources: DTI, PVLog.de and PVOutput.org. Over 1000 site data (measurements from PV systems) were collected from these three sources. Figure 2 shows a histogram of the difference between measured capacity factors and capacity factors simulated with the GSEE.

**Figure 2**: Histogram of the difference between the three modelled and measured capacity factors. Negative values infer underestimates of the capacity factor and positive values overestimates. The blue graph corresponds to modelled data with MERRA, the green graph to MERRA-2 and the red graph to SARAH. (Pfenninger and Staffell, 2016)

The mean of this differences is taken by Pfenninger and Staffell (2016) as basis for a bias correction. They used it to calculate one correction factor for each simulation (MERRA, MERRA-2 and SARAH).

Additionally to the power output simulation of a wind or solar plant at a specific location, renewables.ninja also provides capacity factors per country. Pfenninger and Staffell (2016) have performed randomized (tilt and azimuth angel) national-scale simulations with MERRA and SARAH data to estimate averaged capacity factors per country. Also, these data were analysed against the measured site data. To do so only countries with a least 10 available measurement sites were used. The measured capacity factor per country was then calculated as the mean of all capacity factors per site in one country. Transmission Network Operator (TNO) reports were used as a further data source of the national aggregated capacity factors. Figure 3 shows an example of the results for Great Britain.



**Figure 3:** Capacity factors per season for Great Britain from aggregated measurements site data (solid line, a and b), from simulation with MERRA and SARAH (dotted line, a and b) and from Transmission Network Operator (TNO) reports (c).

Currently they are working on the implementation of ERA5 dataset as a source for the GSEE. The data are projected to be available before the start of this master thesis. Therefore, the newly generated data will be used for our study. As fallback scenario the already available data are will be used.

### Installed PV capacities

The following datasets will be used to gather the installed PV capacities for each country in Europe, which is necessary to calculate the absolute PV power output:

* EurObserv’ER: Consortium that monitors the development of renewable energies in Europe
* IRENA: International Renewable Energy Agency
* ENTSO-E: European Network of Transmission System Operators for Electricity
* Eurostat: European Statistical Office

All of them provide data about current (at least yearly) installed PV capacity data within Europe. These sources were also taken by Pfenninger and Staffell (2016) and we will analyse and further process them during work.

## Method

### Weather regime classification

Our study will use the 500hPa geopotential height and surface downward solar radiation variables from ERA5 in the domain 80°W to 40°E, 30°N to 90°N. The domain is chosen according to Grams *et al.* (2017) for easy comparison, and is also reasonable for our meteorological field investigates since it captures the large scale circulation that affects Europe. We will perform EOF analyses on these two variables and with the capacity factors from the simulation by renewables.ninja. An EOF analyses decomposes a dataset into statistically orthogonal modes that describe the variability of the data. For metrological datasets, a few modes are often sufficient to explain a large fraction of the total variability in the data, which is helpful to assess the key patterns of the variability and to further analyse them. The input parameter for the EOF analysis with the fewest resulting modes, which explain the highest percentage of the variability, will be used to further work with. This is just a first criterion which can be used, the final decision will be taken during the work because it depends on the results and its analysis.

The generated subspace spanned by the leading EOFs will be used to group the data with the k-mean clustering method. It is commonly used to classify weather regime with this method (Michelangeli, Vautard and Legras, 1995; Cassou, 2008) and we will make use of this approach as well. Clustering our radiation and geopotential height EOF analysis results, will yield in weather and radiation regimes. For simplification we will use hereafter just the term weather regime even if it includes the radiation analysis as well. Generally, clustering techniques are used to group data with similar characteristics by minimizing the variance within the clusters. The difficulty lies in the definition of the number of clusters. There are several methods to find the most suitable number of clusters, which will be evaluated during our work. For the Euro-Atlantic region often four clusters are used to define the weather regimes (Michelangeli, Vautard and Legras, 1995; Cassou, 2008; Van Der Wiel *et al.*, 2019). Grams *et al.* (2017) used 7 clusters in his study and therefore ended up with 7 weather regimes. For easy comparison we may use the same number of clusters, but other approaches are feasible and will be evaluated during work with the data. Finally, we will have the information which weather regime prevails on which time period. As soon as we have this attribution, we can assign capacity factors to the corresponding weather regimes for further analysis.

### Capacity factor estimates

We will consider two possible approaches to use the data provided by renewables.ninja. The first is based on national aggregated capacity factors analogous to Grams *et al.* (2017). The advantage of this dataset is the included bias correction described above. But one must consider that this is strongly depended on the amount and position of measurement sites in a country. For example, for Spain they only have 14 available measurement sites that are mostly located at the northern coast. This leads to a statistically unrepresentative sample (Pfenninger and Staffell, 2016). Nevertheless, a further advantage of this approach is that we would be in line with the study of Grams *et al.* (2017) which makes further comparison or combination analyses of wind and solar power output variability easier.

The second approach is based on using the GSEE with the ERA5 dataset to calculate the capacity factors for each ERA5 grid cell by our own. One advantage of this approach is data consistency, since we are doing the weather regime definitions based on the ERA5 dataset as well. But this will also be the case if the implementation of ERA5 into renewables.ninja will be published in time. Furthermore, we can specify the size of the regions for further use in our analysis. In other words, we could replace capacity factor per country with capacity factor per region of interest with only the grid resolution of ERA5 as constrain. Especially, for large countries like Germany, this could be useful since the potential difference between its capacity factors for northern and southern regions could be substantial.

Both approaches include the attribution of the capacity factor to the before defined weather regimes. As described above, this will be done by assigning time periods to a specific weather regime and capacity factors during the same time period will then be the attributed to them. We will use the attributed capacity factors to calculate a mean capacity factor per weather regime, season and region/country. The difference between this mean capacity factors and the mean capacity factors for the whole season of a country/region determines whether the impact of the weather regime exhibits over- or underproduction. Furthermore, this quantity can be normalised by the mean capacity factor for the whole season of a country/region to get a percentage deviation:

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 2 |

where CFwr,country,season is the mean capacity factor of a specific weather regime, country/region and season [unitless] and CFcountry,season is the mean capacity factor of a country/region for the whole season [unitless].

Analogous to Grams *et al.* (2017) we will calculate the European capacity factors per weather regime and season by summing up the capacity factor per weather regime, season and country/region and weighting them per land area:

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 3 |

where CFwr,Europe,season is the mean capacity factor per weather regime and season for Europe [unitless], Acountry is the land area per country [km2] and AEurope is the land area of Europe [km2].

Multiplication of capacity factors with installed capacities yields power output (Eq. 1). This can be used to expand Eq. 3 which gives us the total solar power output of Europe per weather regime and season. And to get the percentage deviation it can be normalised by the mean solar power output of Europe per season:

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 4 |

where ICcountry is the installed capacity per country [W] and PEurope,season is the mean solar power output of Europe per season [W].

We will use Eq. 4 as an expression for the variability. The results will provide a seasonal percentage of over- or underproduction per weather regime within Europe relative to the mean. This will be used to answer research question one. The only unknown we still have is the installed capacity per country. Pfenninger and Staffell (2016) used four data sources to obtain the PV installed capacity per country: EurObserv’ER, IRENA, ENTSO-E, Eurostat. We will analyse them and decide how and which of them or others will be used during our study.

The main research question 2 will also be addressed with Eq. 4. The goal is to reduce the variability as much as possible. Therefore, we need to minimize Eq.4 with ICcountry as unknown. And this for every weather regime with the prioritization of the weather regime which explain the highest variability. Constraints are the already installed capacity per country and the aimed total installed capacity for PV systems in Europe. Further, constraints to address can be a maximum difference of installed capacity per country/region to consider transmission limitation between countries/regions.

# Timeline and milestones

|  |  |  |
| --- | --- | --- |
| **Date / Milestone Nr.** |  | **Task / Milestone** |
| **03.08. – 31.08** |  | Review literature |
| **03.08. – 12.10.** |  | Perform EOF analysis and define weather regimes |
| **MILESTONE I** |  | Decision on the most suitable variable for further analysis, based on the EOF analyse results |
| **12.10. – 01.12.** |  | Attribute and process capacity factors |
| **01.12. – 31.12.** |  | Evaluate current multiday PV power output variability in Europe |
| **MILESTONE Ⅱ** |  | Overview of current situation in Europe |
| **01.01. – 28.02.** |  | Assess future installed PV capacity per country to reduce variability in Europe |
| **MILESTONE Ⅲ** |  | Result for potential future PV installed capacity in Europe / Work on data completed |
| **24.08. – 06.03.** |  | Write master thesis |
| **MILESTONE IV** |  | Submit master thesis |
| **06.03. – 31.03.** |  | Review and corrections |
|  |  |  |

# References

Cassou, C. (2008) ‘Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic Oscillation’, *Nature*, 455(7212), pp. 523–527. doi: 10.1038/nature07286.

Dawson, A. (2016) ‘eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data’, *Journal of Open Research Software*, 4, pp. 4–7. doi: 10.5334/jors.122.

Delucchi, M. A. and Jacobson, M. Z. (2011) ‘Providing all global energy with wind, water, and solar power, Part II: Reliability, system and transmission costs, and policies’, *Energy Policy*. Elsevier, 39(3), pp. 1170–1190. doi: 10.1016/j.enpol.2010.11.045.

Graabak, I. and Korpås, M. (2016) ‘Variability Characteristics of European Wind and Solar Power Resources—A Review’, *Energies*, 9(6), pp. 1–31. doi: 10.3390/en9060449.

Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I. and Wernli, H. (2017) ‘Balancing Europe’s wind-power output through spatial deployment informed by weather regimes’, *Nature Climate Change*, 7(8), pp. 557–562. doi: 10.1038/NCLIMATE3338.

Heide, D., von Bremen, L., Greiner, M., Hoffmann, C., Speckmann, M. and Bofinger, S. (2010) ‘Seasonal optimal mix of wind and solar power in a future, highly renewable Europe’, *Renewable Energy*, 35(11), pp. 2483–2489. doi: 10.1016/j.renene.2010.03.012.

Hennermann, K. and Yang, X. (2018) *ERA5 data documentation*, *European Centre for Medium-Range Weather Forecasts*. Available at: https://confluence.ecmwf.int/display/CKB/ERA5+data+documentation (Accessed: 22 April 2020).

Huld, T., Gottschalg, R., Beyer, H. G. and Topič, M. (2010) ‘Mapping the performance of PV modules, effects of module type and data averaging’, *Solar Energy*, 84(2), pp. 324–338. doi: 10.1016/j.solener.2009.12.002.

Hulme, M. (2016) ‘1.5 °C and climate research after the Paris Agreement’, *Nature Climate Change*. Nature Publishing Group, 6(3), pp. 222–224. doi: 10.1038/nclimate2939.

Jäger-Waldau, A. (2019) *PV Status Report 2019*, *EUR 29938 EN, Publications Office of the European Union*. Luxembourg. doi: 10.2760/326629.

Lauret, P., Boland, J. and Ridley, B. (2013) ‘Bayesian statistical analysis applied to solar radiation modelling’, *Renewable Energy*, 49, pp. 124–127. doi: 10.1016/j.renene.2012.01.049.

Michelangeli, P. A., Vautard, R. and Legras, B. (1995) ‘Weather regimes: recurrence and quasi stationarity’, *Journal of the Atmospheric Sciences*, pp. 1237–1256. doi: 10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

Mills, A. and Wiser, R. (2010) *Implications of Wide-Area Geographic Diversity for Short- Term Variability of Solar Power*, *Energy*. Berkeley, CA (United States). doi: 10.2172/986925.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) ‘Scikit-learn: Machine Learning in {P}ython’, *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Pfenninger, S. and Staffell, I. (2016) ‘Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data’, *Energy*, 114, pp. 1251–1265. doi: 10.1016/j.energy.2016.08.060.

Ram, M., Bogdanov, D., Aghahosseini, A., Oyewo, A., Gulagi, A., Child, M. and Fell, H.-J. (2019) ‘Global Energy System based on 100% Renewable Energy – Power, Heat, Transport and Desalination Sectors’, *Study by Lappeenranta University of Technology and Energy Watch Group*. Berlin, 1(March), pp. 1–321.

Ridley, B., Boland, J. and Lauret, P. (2010) ‘Modelling of diffuse solar fraction with multiple predictors’, *Renewable Energy*, 35(2), pp. 478–483. doi: 10.1016/j.renene.2009.07.018.

Santos-Alamillos, F. J., Pozo-Vázquez, D., Ruiz-Arias, J. A., Von Bremen, L. and Tovar-Pescador, J. (2015) ‘Combining wind farms with concentrating solar plants to provide stable renewable power’, *Renewable Energy*, 76(2015), pp. 539–550. doi: 10.1016/j.renene.2014.11.055.

Stram, B. N. (2016) ‘Key challenges to expanding renewable energy’, *Energy Policy*, 96, pp. 728–734. doi: 10.1016/j.enpol.2016.05.034.

Sweerts, B., Pfenninger, S., Yang, S., Folini, D., van der Zwaan, B. and Wild, M. (2019) ‘Estimation of losses in solar energy production from air pollution in China since 1960 using surface radiation data’, *Nature Energy*, 4(8), pp. 657–663. doi: 10.1038/s41560-019-0412-4.

Van Der Wiel, K., Bloomfield, H. C., Lee, R. W., Stoop, L. P., Blackport, R., Screen, J. A. and Selten, F. M. (2019) ‘The influence of weather regimes on European renewable energy production and demand’, *Environmental Research Letters*, 14(9). doi: 10.1088/1748-9326/ab38d3.

Wild, M., Folini, D., Henschel, F., Fischer, N. and Müller, B. (2015) ‘Projections of long-term changes in solar radiation based on CMIP5 climate models and their influence on energy yields of photovoltaic systems’, *Solar Energy*, 116, pp. 12–24. doi: 10.1016/j.solener.2015.03.039.