GEO4 – 2011 Data Analytics for Sustainability

Assignment ESM Module – Renewable Energy Droughts

Due date: Fri., Jan. 23, 2026 at 23:59 hrs

Background and logistics

The most obvious measure that contributes to the decarbonization of our energy supply system is to increase the penetration of renewable energy installations, such as wind turbines and solar PV panels. These technologies are mature and cost-competitive with the traditional, fossil-fuel based energy conversion methods. However, due to their intermittent nature, solar and wind energy come with challenges related to balancing supply and demand of electricity on various time frames, and energy security concerns at the national/regional level. In order to come up with mitigation measures, the first step is to thoroughly analyse the nature of this intermittency, and to understand its temporal and spatial properties. You will work in groups of two students on an assignment where you learn how to do just that. For background, you may consult the reference list [1-3] at the end of this document.

Should you choose the ESM module, this assignment will count for 65% of your grade for the DAFS course. We estimate a workload of about 60-70 hours to complete this assignment (excluding the time to go through the 3 tutorials), so make sure you start on time. *Please upload your Jupyter notebook to BrightSpace by the due date specified above.* Prior to that, please inform the module instructor prof. M. Gibescu as to the region of your choice for Exercise 4, as well as the additional dataset(s) and open research question that you plan to answer for the last part of the assignment, which counts for 20% of the ESM module grade. This can be done by completing the *sign-up sheet at the latest by week 6 of the course (link will be sent via BrightSpace announcement).*

Note that this year both pre-set exercises 1-6 will be done in groups, as well as the *open research question*. Please specify each students' individual contributions in your report. When writing your report, please keep in mind the rubric provided as an appendix at the end of this assignment.

Good luck!

Overview: In this final assignment for the Energy System Modelling module, you will first look into the concept of Renewable Energy Droughts and then you will have the opportunity to pose and answer an energy system related research question of your choosing. You will use the skills you learned in the previous tutorials, but you will also need to use packages/functions that have not been introduced in the tutorials. Hence, you might need to look up the documentation of relevant Python packages, and use Google/Stack-overflow/ChatGPT to find solutions to the problems you will encounter. *Follow the steps below as a reference, note that these are identical to the instructions provided in the Jupyter notebook.*

Exercise 1 - ERA5 data download and merge (0,5/10 pts)

In this assignment, we will again use the [ERA5 Reanalysis dataset] and a vector map of NUTS regions. You can use the same NUTS map that you used in the tutorials. For the ERA5 dataset, we provide you with a file including a six-month period (1 July 2022 - 31 Dec 2022) with a three-hourly time resolution. This dataset contains the variables 'ssrd', 'u100', 'v100' and 't2m' for mainland Europe, using the following coordinates:

North = 71.22 South = 34.31 East = 51.29 West = -23.62

The above variables are needed to later create time series for solar and wind energy production. Your first task is to complement this dataset by downloading the data for the period between 1 January 2022 - 30 June 2022. Consider the same variable sets, time resolution and coordinates as with the provided data. After downloading the data, merge the downloaded ERA5 dataset with the provided ERA5 dataset to create one single xarray array, containing data for all days of 2022. Please provide the code you have used as well as a concise summary of the data set variables and attributes using xarray.Dataset.info.

Exercise 2 - Determining solar and wind power capacity factors (0.5/10 pts)

The weather data in the ERA5 dataset can be used to estimate the capacity factors for wind turbines and solar photovoltaic (PV) panels. The capacity factor be defined as follows:
"*The ratio of the electrical energy produced by a generating unit for the period of time considered to the electrical energy that could have been produced at continuous full power operation during the same period*." (EIA)

The capacity factor of a PV panel can be estimated using the downwards surface solar radiation and the temperature. Using the data from the ERA5 dataset, we will add the variable 'cf_pv' to the era5 xarray array, which shows the capacity factor for each grid cell and for each 3-hour timestep. To do so, you need to use a function, based on the supplementary material of Brown, Farnham and Caldeira (2021) [2] (see Jupyter notebook for the actual code). Next, we use the wind-related variables 'u100' and 'v100' in the ERA5 dataset as inputs to determine the capacity factor for wind turbines. First, we need to compute the absolute total wind speed from these 2 variables in the ERA5 dataset. If you have questions about the exact meaning of these variables, please have a look of the metadata of the ERA5 Reanalysis dataset. Save this new variable in the era5 xarray array using the variable name 'ws'. To determine the capacity factor of wind turbines, we consider a simplified power curve of a wind turbine. A power curve shows the power output of a wind turbine at specific wind speeds. You will use the following power curve in your analysis:

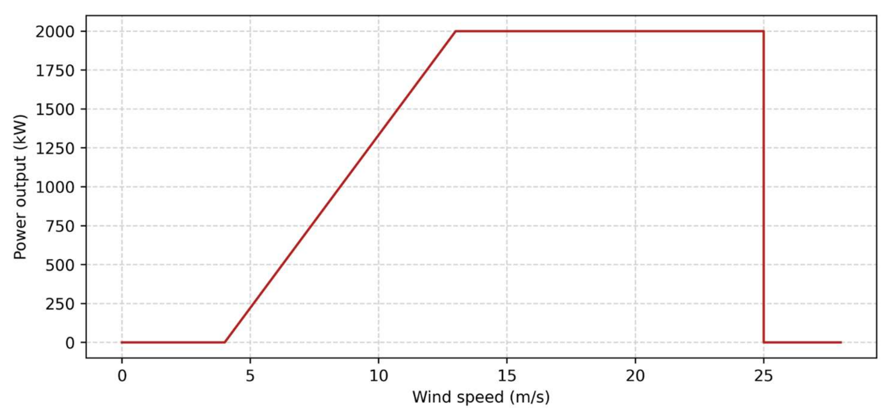

Fig. 1 - Simplified speed to power conversion curve for a wind turbine of 2 MW rated power.

The considered turbine has a cut-in wind speed of 4 m/s, a rated wind speed of 13 m/s, a cut-out wind speed of 25 m/s and a rated capacity of 2 MW. This means that at wind speeds below 4 m/s and above 25 m/s, there is no power output. Between 4 and 13 m/s, we assume for simplicity that the power output increases linearly from 0 to its rated capacity. At wind speeds between 13 m/s and 25 m/s, the power output equals the rated capacity.

Your next task will be to add a new variable 'cf_wind' in the era5 xarray array, portraying the capacity factor for wind turbines for each grid cell and each 3-hour timestep using the above power curve. To do this, create a function to convert the wind speed in the era5 xarray array to wind power. Some tips:

1. Take a look at the provided function to determine the capacity factors for solar PV systems, and use it as an inspiration to create your own function.
2. Consider to use the [xarray.DataArray.where]() function.
3. Do not forget to convert the wind power from physical units to the a-dimensional capacity factor
4. Give it a try yourself before relying on AI to help develop this function!
5. Perform some sanity checks (Do individual values make physical and technical sense?) to see if your function works properly.

Exercise 3 - Plotting capacity factors for NUTS-1 regions (0.5/10 pts)

Your next exercise will be to plot the average capacity factor for both solar PV systems and wind energy installations of every NUTS-1 region in Europe. Import the [database with NUTS regions]() and determine the average capacity factor per region for each of the 2 technologies. Present the results using choropleth maps. Take care of the layout of the figure (readability of ticks, axis labels, figure titles, units, legend(s), comparability of figures, etc.). Use the subplot function to plot these two maps in one figure using two subplots side by side. Comment on the data analytic insights that can be gained from the figure.

Exercise 4 - Comparing solar PV and wind profiles of different regions (3/10 pts)

In the fourth exercise of this assignment, we want to compare the wind power and solar PV power profiles over time between the different NUTS-1 regions.

As a first step, we need to determine the average wind and solar capacity factors for every NUTS-1 region for every 3-hour timestep of the year. To do this, it is recommended to create a pandas dataframe with every timestep as the index, and the specific NUTS-1 region as a column name.

Note: *Running this script will take long. Hence, it is highly recommended to save the dataframes you created so you can read them later from a file, without having to re-run the script.*

Please provide a detailed explanation of how you approached this, the developed code as well as a summary of the created pandas data frame (number of lines, number of columns and average values per column; are these in the same range as in exercise 3?). Explain how you checked that your results are accurate.

As a second step, you will deep-dive into one specific NUTS-1 region (not located in the Netherlands and not the same one of your peers). Create a bar plot showing the average capacity factor for every month for wind turbines and PV panels for your selected region. Present the capacity factors for both technologies in one figure. You can use the dataframe created above for this exercise, for instance using the *resample*-function. Please provide the developed code and an explanation of your figure. Do the values presented in your figure make physical/technical sense? Justify your statements.

As a third step, using the pandas dataframe with the average wind and PV capacity factors per time step and region created under step 1, determine the correlation of the **3-hourly values** of the two considered capacity factors between your selected NUTS-1 region and all other NUTS-1 regions. Plot the results.

Please provide the developed code and choropleth maps showing the correlations for the PV capacity factor and another one showing the correlations for the wind power capacity factor.

Note: *You will be required to use Python functions that we did not introduce before in this course! Use Google and the Python library documentation to find relevant functions.*

Q4.1. [1 point] In the cell below, shortly discuss your observations based on the plots you created. What trends do you observe, and what causes these trends?

Subsequently, repeat the third step using the **weekly** average values for every NUTS-1 region, instead of for the 3-hourly values. Again, plot your results for the solar PV and wind correlations between your selected NUTS-1 region and all other regions in Europe.

Please provide the developed code and the choropleth maps also in this case.

Q4.2. [1 point] What differences do you observe between the plots based on the 3-hourly values and weekly average values? Discuss the main findings, and explain what causes them.

Q4.3. [1 point] Explain why graphs showing the correlations in wind and solar energy production patterns are relevant for policy makers to provide insights into energy system design in case of high penetration of renewable energy.

Exercise 5 - Modelling RES droughts (2/10 pts)

One of the main challenges for a fully-renewable based energy system is the concept of renewable energy droughts (often referred to as *Dunkelflaute* [3], *Dark Doldrums or RES Drought*), which are moments at which there is simultaneously no or little solar **and** wind power generation.

In this exercise, we are going to identify the share of time with a RES drought for every NUTS-1 region in mainland Europe, starting with the dataframe of the 3-hourly solar and wind generation per NUTS-1 region you created earlier. To continue, you can assume either one of the following two simplified definitions for a period with a renewable energy drought for a specific region:

- Definition according to [2]: A time period during which **both** capacity factors for wind turbines and PV panels are in the **lowest 10% quantile in the considered dataset** for the considered NUTS-1 region, or
- Definition according to [3]: A time period during which **both** capacity factors for wind turbines and PV panels are **below 10% of their respective nominal capacity**.

Specify which definition you chose and justify your choice.

Determine the share of time with a RES drought in the considered dataset for each region for both a 3-hourly timeframe and a weekly timeframe. Plot the results.

Please provide the developed code and separate maps for the 3-hourly values and for the weekly values.

Tip: *To perform this analysis, it is recommended to create a new dataframe with every timestep as the index, and the different NUTS-1 regions as the columns. If a time period for a specific region experience a renewable energy drought according to the definition above, mark it with 1. Otherwise, mark it with 0.*

Q5.1. [1 point] Discuss the trends you observe in your plots. How does the difference in 3-hourly and weekly RES drought values affect the way policymakers should design their energy systems to ensure adequacy, i.e. electricity demand can be served at all times?

Q5.2. [1 point] Critically reflect on the presented approach to identify renewable energy droughts for NUTS-1 regions. Discuss 3-5 main shortcomings of this approach, and their impact.

Exercise 6 - Clustering RES droughts (1.5 points)

When designing our future renewable energy system for the whole Europe, policymakers will be interested in the similarities and differences in the RES drought patterns between regions. Neighbouring regions in Europe are linked through interconnectors, which are used to transport electricity between the different regions. Hence, in case of a renewable energy drought, energy system adequacy may be ensured through the import of electricity from other regions.

In this exercise, we ask you to group the different NUTS-1 regions into a number of clusters with similar renewable energy drought patterns. This can be done by a method called *clustering*, in which the NUTS-1 regions can be subdivided into a predefined number of clusters with similar properties. You will need the dataframe with renewable energy drought data created in Exercise 5 to do this. Use the **3-hourly values** to create clusters.

Different clustering methods exist. Look up relevant clustering methods in the literature and find the documentation of a Python package supporting your preferred clustering method. Experiment with a number of clusters between 4-6 until the results are interpretable from an energy system perspective. Present your end result using a map displaying the different clusters.

Note: *A commonly-used Python package for clustering is sklearn. Google for the appropriate documentation to conduct the analysis.*

Q6.1. [0,5 points] Shortly describe what clustering method you used, why you selected this particular clustering method and how this method works to create its clusters.

Q6.2. [1 point] What trends do you observe? Describe how policy makers could use such figures to promote an adequate European energy system design in case of high renewable energy penetration, in particular regarding investments for interconnector capacity.

Open research question (2/10 pts)

Extend the analysis on capacity factors, correlations and renewable energy droughts, for instance by eliminating some of the shortcomings you identified in Q5.2. Or propose a new energy system related question that you are interested in. Relations with other sectors of the economy, social science aspects, land use are also interesting to explore. Feel free to download extra data from CDS, NEWA (New European Wind Atlas), the International Energy Agency IEA Europe Data explorer, or other publicly available sources to conduct your analysis. Other suggestions include electricity demand and price data from the ENTSO-e Transparency platform, interconnection capacities from TYNDP, installed capacities of wind energy in Europe from https://zenodo.org/records/7558885 etc. Bring your creativity here, and have some fun with this open research question!
Describe the data you used, the steps you have taken for your analysis, and present your results with attractive plots. Also mention clearly who worked on which of the sub-questions in your open RQ.

References

[1] L. Ramirez Camargo, M.L. Lode, T. Coosemans, "Assessing the relevance of renewable energy resources availability for the existence of Energy Cooperatives in Europe," *Energy Proceedings*, Vol. 29, 2022, https://doi.org/10.46855/energy-proceedings-10327.

[2] P.T. Brown, D.J. Farnham, K. Caldeira, "Meteorology and climatology of historical weekly wind and solar power resource droughts over western North America in ERA5," *Springer Nature Applied Sciences* **3**, Article no. 814, 2021. https://doi.org/10.1007/s42452-021-04794-z.

[3] B. Li, S. Basu, S. J. Watson, H. W. J. Russchenberg, "Mesoscale modeling of a Dunkelflaute event," *Wind Energy*, Volume24, No. 1, Jan. 2021, pp. 5-23, https://doi.org/10.1002/we.2554.

Appendix: Rubric for grading the ESM module

	Student names:				Weight	Grade
Pass/Fail Items						
No.			Pass	Fail		
1	**Download data using the CDS API**	The student is able to retrieve data from the CDS by adapting the provided code from tutorials to do so.			0,25	
1	**Merging data**	The student can organize the data in a way that it is useful for further processing.			0,25	
Scale items						
No.		Insufficient (< 5.5)	Acceptable (5.5-7)	Good (> 7)		
2 & 3	**Determining and plotting average capacity factors per NUTS-1 region**	The student is unable to complete the code to compute the capacity factors for Wind power. Or, the student produces maps of CF for PV and Wind power but the ranges/averages per region are not correct.	The student produces maps of CF for PV and Wind power averaging per NUTS-1 region and the ranges/averages per region are correct. The maps include adequate legends and labels.	Additionaly to generating the maps as decribed under "Acceptable", the student provides a thorough "sanity check". Excellent work includes the checks of time series for individual regions and comparison of average values with values found in literature.	1	
4 part 1	**Creating solar PV and wind CF time series for all NUTS-1 regions**	The student is unable to complete the code to calculate the CF per NUTS-1 region and time step. Or, the provided code results in a dataframe with the wrong dimensions and/or innacurate values.	Student develops code able to calculate the CF per NUTS-1 region and time step, the resulting dataframe has the right dimensions and some minimum explanation is provided about the accuracy of the values in the dataframe.	Student develops code able to calculate the CF per NUTS-1 region and time step, the resulting dataframe has the right dimensions and a solid quality check of the accuracy of the values is performed. Excellent work includes an analysis and comparison with the results obtained in exercises 2-3.	1	
4 part 2	**Time series resampling and plotting for one selected region**	The student is unable to complete the code. Or, the code runs but the figure is of the wrong type, contains incorrect values and/or is missing appropriate labeling and legend. No explanation is provided.	The correct figure is provided, including appropriate labeling and a legend. The presented values seem feasible. A minimum of explanation is provided but is not very convincing.	The figure is provided as described under "Acceptable". The presented values are feasible and the student presents arguments about seasonality and location of the NUTS-1 region to justify the presented results.	1	
4 part 3	**Correlations between the selected NUTS-1 region and the rest of the regions in Europe, calculation and plotting**	The student is unable to complete the code. Or, the student produces correlation maps but the values are incorrect and/or are missing appropriate labels and legends. Not all questions Q4.1-Q4.3 are tackled, or those that are tackled provide incorrect arguments.	Code and maps are provided. The map has proper labels and legends and the values seem feasible. Discussion questions Q4.1-Q4.3 are all answered but some of them in a superficial way.	Code and maps are provided. The map has proper labels and legends and the values are feasible. Solid arguments are provided for Q4.1-Q4.3 . Excellent work includes supporting the arguments with reference to scientific literature.	1	
5	**Renewable energy droughts**	The student is unable to complete the code. Or, the student produces the RES drought maps but the values for either the 3-hourly or weekly droughts are incorrect and/or maps are missing appropriate labels and legends. Not all questions Q5.1-Q5.2 are tackled, or they provide incorrect arguments.	Code and maps are provided. The maps include proper labels and legends. The ranges of both the 3-hourly and weekly RES droughts are correct. Questions Q5.1-Q5.2 are both addressed but some of them in a superficial way.	Code and maps are provided. The maps include proper labels and legends. The ranges of both the 3-hourly and weekly RES droughts are correct. Q5.1 and Q5.2 are properly addressed, with good argumentation. Excellent work would reflect on at least five different shortcomings for Q5.2, with support from the literature.	2	
6	**Clustering RES droughts**	The student is unable to complete the code for clustering. Or the code produces a clustering result and associated map, but appropriate labeling and legend is missing. Q6.1 answer is missing or shows lack of understanding of the clustering method selected.	Code and map are provided. The map includes proper labeling and legend. Q6.1 and Q6.2 are both addressed but some of them in a superficial way.	Code and map are provided. The map includes proper labeling and legend. Q6.1 and Q6.2 are both addressed, with good argumentation. Excellent work would support the answer to Q6.2 with references to scientific literature.	1,5	
7	**Open research question**	The student is unable to complete the code to answer an energy system related research question of their own choosing. Or, the code runs but inappropriate/incomplete data sets are used or for whatever reason the results do not make sense.	Code and appropriate visualizations are provided to answer the proposed research question. Figures include proper labeling and legends. Some discussion of the results is given, but is superficial or unconvincing in places.	Code and appropriate visualizations are provided to answer the proposed research question. Figures include proper labeling and legends. Discussion of the results is solid and provides good argumentation and reflection. Excellent work supports the answer to the open RQ with references to scientific literature.	2	
	Final grade				**10**	