Statistical Modeling of Widespread Energy Droughts

Alejandro Alarcon Gonzalez

June 2025

1 Introduction

The electricity generated with photovoltaic panels (PV's) and wind turbines (WT's) is subject to ever-changing atmospheric conditions and therefore, highly variable. On the other hand, at the electrical grid level, the energy supply must always be in balance with demand. Improving cross-border interconnections between European electricity grids enhances the security of electricity supply. To this end, the EU has set an interconnection target of at least 15% by 2030 (see [7]). Given that European electricity systems increasingly rely on weather-dependent renewable energy sources, such interconnections offer a strategic solution to address the challenge of multiday fluctuations in renewable energy production. The surplus energy generated in one country can be transmitted to neighboring countries experiencing higher demand, thereby improving overall system stability. However, greater interconnection may also increase system vulnerability during extreme weather events that simultaneously affect multiple regions, potentially amplifying supply disruptions.

Hence, it becomes important to estimate the occurrence of wide-spread *energy droughts* or *dunkelflautes*, which are prolonged periods of low wind and solar energy (see Duchêne et al., [5]). In this study, we consider energy droughts whose geographical extension crosses European country boundaries. We do so because a number of recent studies (see Brown et al., [1]; Millstein et al., [11]; Liu et al., [10]) have stressed the link between large-scale weather patterns and widespread energy droughts.

Estimating the spatial extent and occurrence of such dunkelflautes can help assess the vulnerability of an interconnected grid. Due to the pressing need for an energy transition, the scientific community has been turning its attention to dunkelflautes during the last years. For the Belgian case, Duchêne et al. found that relying solely on either solar or wind energy is very ineffective in mitigating dunkelflautes.

The scientific tool proposed to carry out the estimation of dunkelflautes is an extreme value analysis. With such analysis, it is possible to quantify the probabilistic behavior of a process at unusually high (or low) levels (see Coles, [2]; de Haan and Ferreira, [9]). In the case covered by this document,

the process consists of the time evolution across a subregion of Europe of the energy production from WT and PV combined. The branch of statistics that supports the proposed analysis is called extreme value theory (EVT). There are three main results in this study. These are: i) the estimation of energy droughts in Ukkel, ii) the estimation of the spatial extent of energy droughts for a subregion of Western Europe and iii) the probability estimation (in terms of return periods) of energy droughts for the regions derived by ii).

2 Data

The data analyzed in this document consists of *capacity factors* of solar photovoltaic power generation (SPV) and onshore wind power generation (WON). The capacity factor (CF) refers to the ratio of actual generation to installed generation capacity. The CF is then a dimensionless quantity. The dataset (retrieved from [4]) is spatiotemporal and is computed from reanalysis data. It spans the period from 1979 to the present and covers the whole of Europe. We consider the data at a daily resolution, since this is the temporal scale of the energy drop events addressed in this study. The horizontal resolution is given by a grid with 0.25° (roughly 31 km) spacing.

This study focuses on the subregion of Western Europe bounded by 4° E to 10°E and 47°N to 52°N (see Figure 1). It consists of 525 gridpoints. The guiding principles to choose this region are that i) estimating the occurrence of dunkelflautes taking place across country boundaries can help assess the vulnerability of an interconnected grid, and ii) it is a large West-European region that is not crossed by major mountain ranges. In turn, the reason why this last criterium is interesting is two-fold: Western Europe is the European region with the highest interconnectivity rates (see the internal report in [7]). In addition, as mentioned in the introduction, energy droughts are linked with large-scale weather patterns, which in turn can be affected by the interaction with the topography of mountain ranges.

In this document we follow Duchêne et al. to represent dunkelflautes by the mix of solar and wind energy

$$CF_{tot} = (1-\delta)CF_{WON} + \delta CF_{SPV}, \quad \delta \in [0,1]. \tag{1}$$

All data analysis in this study is made for CF_{tot}.

3 Methods

It is worth mentioning that EVT, in general, addresses extremely large values, but since this study focuses on extremely low values, the data will necessarily be transformed before proceeding with the different extreme value analysis covered in this document. Such transformations are specific to each analysis, and will be explained in this section.

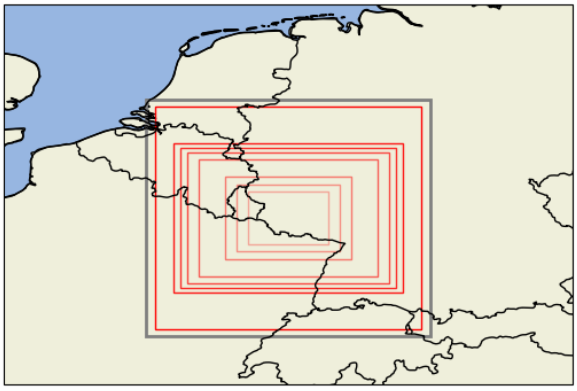


Figure 1: The gray rectangle contains all the grid points that we analyse in this study. The nested regions in red show the spatial extent of very low energy production events for different weights of solar and wind energy in the mix. When only wind energy is considered, the spatial extent is the narrowest (region with lightest red shade). The extent increases monotonically with the weight of CF_{SPV} till this accounts for 70% of the mix (represented by the rectangle with the darkest red), and then remains unchanged for $\delta = 0.8, 0.9, 1$.

3.1 Univariate analysis

The spatiotemporal data can most simply be analyzed by fixing a location from the grid and fitting a statistical model on its temporal data according to EVT. Let us denote the daily capacity factor values by the discrete-time sequence $(X_t^{orig})_{t=1}^m$.

To model the extreme lows of $(X_t^{orig})_{t=1}^m$ we need i) to perform a data transformation such that extremely low values are converted to extremely large ones. With these transformed data, we can use the peaks over threshold (POT) method (see [2, 6]), which focuses on all exceedances of the transformed data over a high threshold u (in this context, the terms exceedances and extremes become interchangeable). Condition ii) for using the POT method (see [6]) is that the parent distribution F (of the data to be modeled) should tend to a Pareto tail, i.e., $1 - F \sim x^{-\alpha}$, for large x values, where $\alpha > 0$ is the tail index. By visually evaluating (with a density plot) different transformations of $(X_t^{orig})_{t=1}^m$, we have found that conditions i) and ii) can be satisfied by the transformed data $X_t = (1/X_t^{orig})^{1/3}$, for $t = 1, \ldots, m$.

We note that the method assumes exceedances to be independent, but in reality, we will likely observe a high capacity factor of WT on a given day to be accompanied by a high value on the next day(s). The same could be said for the

PT case. This is where the notion of *clusters* comes into play (see [2]). Using a well-informed rule to define clusters of exceedances, the clustered values can be identified and replaced by their maximum. The exceedances in the resulting time series are assumed to be independent from each other (extreme events are close to be independent at times that are far enough apart, see [2]). In this study, the clusters of exceedances are defined in terms of the number of days between two synoptic events, which we take to be seven. Let us considered the declustered data $(X_t)_{t\in I}$ (where $I \subset \{1,\ldots,m\}$) to be a sequence of realizations of the random variable X. The conditional exceedances in the declustered time series are then modeled by a generalized Pareto distribution (GPD), with scale parameter σ and shape parameter ξ. That is,

$$\Pr(X > x \mid X > u) \sim \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}, \tag{2}$$

for $1 + \xi(x-u)/\sigma > 0$ and $x > u$, where $\sigma > 0$.

In this document, the parameters of the distribution (2) are estimated with the `R` package `ismev`, which performs a MLE. However, it is more illustrative to present the model fitting in terms of a plot of return levels against return periods. An expression relating these two can be found in [2].

3.2 Spatial Dependence

To estimate the spatial extent of dunkelflautes, we consider one spatial dependence measure from the literature: the *F-madogram*(see [3, 8]).

The F-madogram for locations r_1 and r_2 separated by $h = \|r_1 - r_2\|$ is defined as:

$$\nu_F(h) = \frac{1}{2}\mathbb{E}\left[|F\{Z(r_1)\} - F\{Z(r_2)\}|\right], \tag{3}$$

where $Z(r)$ represents yearly minima of CF_{tot} at location r, and F is its cumulative distribution function. Notably, $\nu_F(h) = 1/6$ for independent year minima and $0 \leq \nu_F(h) < 1/6$ when dependent [3].

The R-package `SpatialExtremes` was used to compute the F-madogram. We note that pairwise distances are not uniformly distributed. To represent the spatial dependence measure against pairwise distances, we define histogram bins with respect to quantiles. For example, by dividing the distance range into ten bins, we use the deciles (10%, 20%,...,100%) as the bin breakpoints. This ensures that the bins contain approximately the same number of pairwise distances. The mean F-madogram is then computed per bin and this process is repeated for $\delta \in \{0, 0.1, \ldots, 1\}$.

As a criterium to estimate the spatial extent of a dunkelflaute, we consider the bin of the shortest grid distances such that the average F-madogram exceeds 0.15 (corresponding to 90% of the independence threshold (1/6)). Then, for each δ, the corresponding bin center was selected as the spatial extent.

3.3 Return period estimation

For return period analysis, we follow Van de Vyver [12], where a method to generate high-dimensional spatial extremes is developed. We denote the spatiotemporal data at locations $r_1, \ldots, r_d$ as $((x_{t,1}^{orig}, \ldots, x_{t,d}^{orig}))_{t=1}^{m}$, and transform them so that extremely low values become extremely large and the data have Pareto margins with tail-index $\alpha = 2$. The exact transformation is given in Appendix B. Let us consider the transformed data as a sequence of realizations of the spatial random process $X = (X(r_1), \ldots, X(r_d))$. An *extreme field* X in the transformed data is defined such that

$$R = \sum_{i=1}^{d} X(r_i) > r_O, \tag{4}$$

where r_O is the 1st percentile of the distribution of the spatial sum R. An example of one extreme field of CF_{SPV} over the studied region on one day is shown in figure 2. To represent the notion of an energy drought we consider declustered extreme fields, analogously to the univariate case, by taking seven days between two synoptic events.

For each region obtained from the spatial dependence analysis, we generate approximately 10^6 such extreme fields, following the algorithm in [12]. The return periods are then estimated for the cases corresponding to 70%, 80%, 90% and 100% of regional gridpoints simultaneously exceeding the marginal 99% percentile.

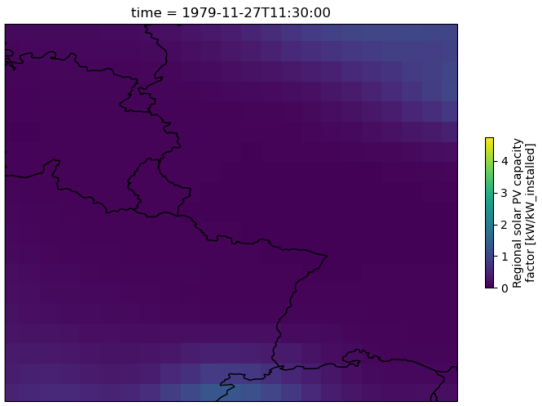


Figure 2: Solar Photovoltaic Capacity on 27 November 1979. The plot illustrates an extremely low combined regional value (see the extreme field definition (4)) of CF_{SPV}.

4 Results and discussion

The generalized Pareto distribution (2) was fitted to the transformed declustered Ukkel data. At a threshold level given by the 99th percentile, the case when only wind and only sun is considered in the energy mix (1) consists of 116 and 100 exceedances, respectively. The model goodness-of-fit was assessed by inspection of probability plots and quantile plots using `ismev`. These plots are shown in Appendix A. The resulting estimation of a (theoretical) local energy production drop is illustrated in figure 3. The return levels shown are obtained by the back transformation of the levels resulting from the POT analysis, so as to represent extremely low levels of the original data $(X_t^{orig})_{t=1}^m$.

We observe from figure 3 that combining SPV with WT at the local level yields better results than a unique energy sourcing from either of the two: the return level for only wind ($\delta = 0$) is between 77 and 132 times lower than the ones corresponding to $\delta = 0.25, 0.5, 0.75$ for a return period of 10 years. For return periods of 20 and 30 years, these proportions only increase. This suggests that a very long period of time has to be waited until the mixed energy sourcing experiences the dramatically low levels observed for the case of only wind very early on. A similar, although more conservative conclusion can be derived from the comparison of only sun with a mixed energy sourcing.

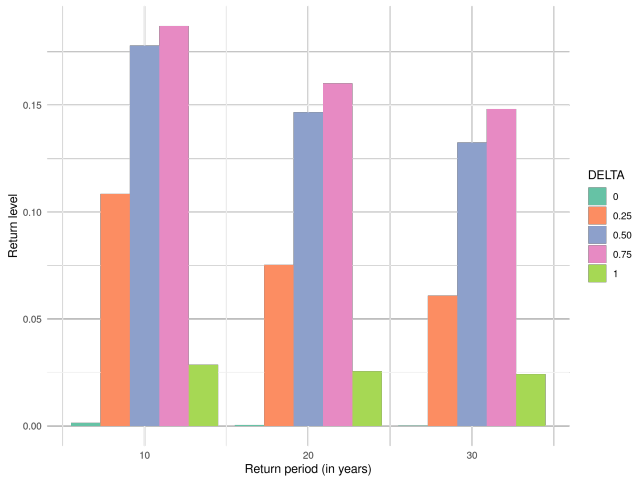
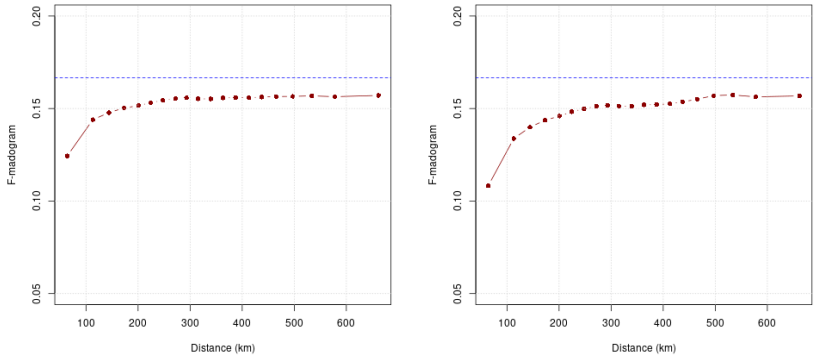


Figure 3: Plots of return levels vs return periods are given for different combinations of the energy mix (1) on the data for Ukkel, Belgium. The return periods studied were 10, 20 and 30 years.



(a) $\delta = 0$ (b) $\delta = 0.2$

(c) $\delta = 0.4$ (d) $\delta = 0.6$

(e) $\delta = 0.8$ (f) $\delta = 1$

Figure 4: F-madograms for the total capacity factor mix (1) over the region of interest in Western Europe. The value of δ is being varied so as to illustrate the effect the weight from either CF_{WON} or CF_{SPV} has on the spatial dependence. The independence line $\nu_F(h) = 1/6$ is plotted in blue.

Figure 4 shows that the breadth of spatial dependence (according to the F-madogram) changes with δ. When only wind energy is considered ($\delta = 0$), spatial dependence is high only within short distances. The more SPV is added to the mix, the farther reaching the dependence becomes, indicating broader spatial extents of dunkelflautes. Therefore, the use of EVT on the data for the studied region (the gridpoints within the gray rectangle in 1) indicates that yearly minima of the wind speed capacity factor have a very limited spatial reach compared to the analogue for solar radiation. Figure 5 shows this evolution, where it can be more clearly observed how the spatial extent of energy droughts increases with δ, reaching a maximum range for high SPV weights (between 70% and 100% of the mix). It is worth mentioning here a source of uncertainty: the estimation of the spatial extent varies when the number of bins changes. In Appendix C, different spatial extent diagrams are observed, for varying number of distance bins. We have chosen the case where the pairwise distances are grouped in 20 bins, for being the most conservative case.

For each δ in the energy mix (1), the geographical extent of an energy drought is illustrated in figure 1. Each of the nested regions is inscribed in a circle whose diameter is given by the spatial extent analysis. The length of the diagonal of these rectangles increases monotonically with the share of solar energy.

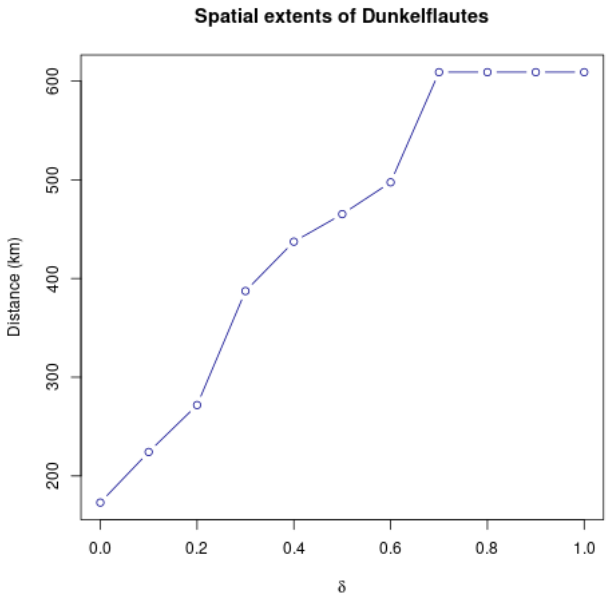


Figure 5: The spatial extent of energy droughts against combinations of the energy mix (1) ($\delta \in \{0, 0.1, \ldots, 1\}$). The extent is calculated as the shortest bin center distance at which the average F-madogram exceeds 90% of the independence threshold (1/6).

The plot in figure 6 shows the estimated return periods of dunkelflautes. For each δ value in the energy mix (1), the estimation was carried out for the corresponding region (see figure 1). The joint exceedence stands for the proportion of gridpoints simultaneously exceeding the marginal 99%-quantile.

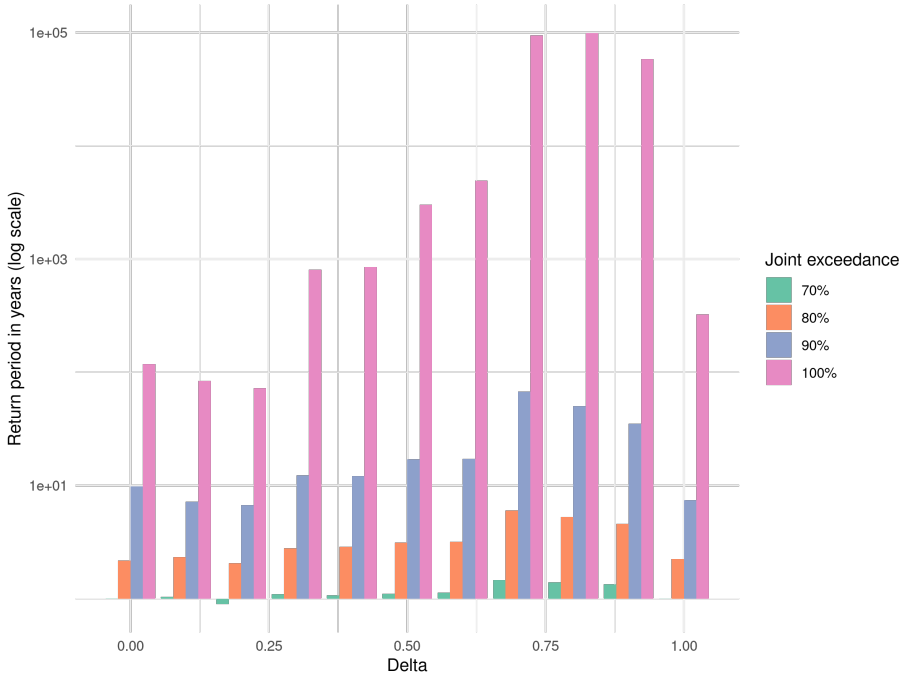


Figure 6: Estimated return periods of dunkelflautes. For each combination of the energy mix (1) there is a typical spatial extent. This was considered in the return period analysis. The joint exceedence stands for the proportion of gridpoints simultaneously exceeding the marginal 99%-percentile. The estimation is derived from a set of approximately 10^6 extreme fields.

We conclude from figures 1 and 6 that although the spatial extent of low energy production from only wind is narrower, the (estimated) temporal frequency of such events is much higher than when solar energy is also considered in the energy production mix. The least frequent production drops were found when solar energy accounts for 70% to 90% of the energy input. This indicates that a mixed use of renewable sources lowers the likelihood of energy droughts in the areas studied. This finding aligns with Duchêne et al. [5], who advise against relying on a single renewable source.

5 Conclusion

We studied the spatial extent of energy droughts based on historical data from 1979 to the present. The energy mix considered was given in terms of the wind and solar capacity factors. By varying their weights in the mix, we showed that the spatial extent of low-production events increases with a higher share of solar energy.

This study also demonstrates the value of combining wind and solar energy in mitigating the risk of energy droughts, in a region of Western Europe. With respect to the historical data and using extreme value theory, we found that wind-only shortfalls are geographically very limited but occur more frequently. In contrast, increasing the solar share in the energy mix significantly broadens the spatial extent of low-production events, yet makes such events considerably rarer. The lowest estimated frequencies of joint production shortfalls were observed when solar energy contributed between 70% and 90% of the energy mix. These results highlight the importance of a diversified renewable portfolio to increase energy resilience at regional scales.

The `R` and `Python` code for reproducing the data analysis, Monte Carlo-based estimations, and visualizations of this study can be found in `https://github.com/jalarcon-ale/Dunkelflaute_estimation`.

References

[1] Patrick Brown, David Farnham, and Ken Caldeira. "Meteorology and climatology of historical weekly wind and solar power resource droughts over western North America in ERA5". In: *SN Applied Sciences* 3 (Oct. 2021). DOI: 10.1007/s42452-021-04794-z.

[2] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer, 2001. ISBN: 978-1-85233-459-8. DOI: 10.1007/978-1-4471-3675-0.

[3] Dan Cooley, Philippe Naveau, and Paul Poncet. "Variograms for spatial max-stable random fields". In: *Dependence in Probability and Statistics*. Ed. by Patrice Bertail, Philippe Soulier, and Paul Doukhan. New York, NY: Springer New York, 2006, pp. 373–390. ISBN: 978-0-387-36062-1. DOI: 10.1007/0-387-36062-X_17. URL: https://doi.org/10.1007/0-387-36062-X_17.

[4] Copernicus Climate Change Service. *Climate and energy indicators for Europe from 1979 to present derived from reanalysis*. 2020. URL: https://cds.climate.copernicus.eu/datasets/sis-energy-derived-reanalysis?tab=overview.

[5] François Duchêne et al. “Current-Day and Future Dunkelflaute Risks for Belgium”. In: *Journal of Applied Meteorology and Climatology* 63.11 (2024), pp. 1427–1440. DOI: 10.1175/JAMC-D-24-0007.1. URL: https://journals.ametsoc.org/view/journals/apme/63/11/JAMC-D-24-0007.1.xml.

[6] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events: For Insurance and Finance*. Vol. 33. Applications of Mathematics. Berlin, Heidelberg: Springer, 1997. ISBN: 978-3-540-60931-5. DOI: 10.1007/978-3-642-33483-2.

[7] European Comission. *Electricity interconnection targets*. 2017. URL: https://energy.ec.europa.eu/topics/infrastructure/electricity-interconnection-targets_en.

[8] A. Gobin and H. Van de Vyver. “Spatio-temporal variability of dry and wet spells and their influence on crop yields”. In: *Agricultural and Forest Meteorology* 308-309 (2021), p. 108565. ISSN: 0168-1923. DOI: https://doi.org/10.1016/j.agrformet.2021.108565. URL: https://www.sciencedirect.com/science/article/pii/S0168192321002495.

[9] Laurens de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. New York: Springer Science+Business Media, 2006. ISBN: 978-0-387-23946-0. DOI: 10.1007/0-387-34471-3.

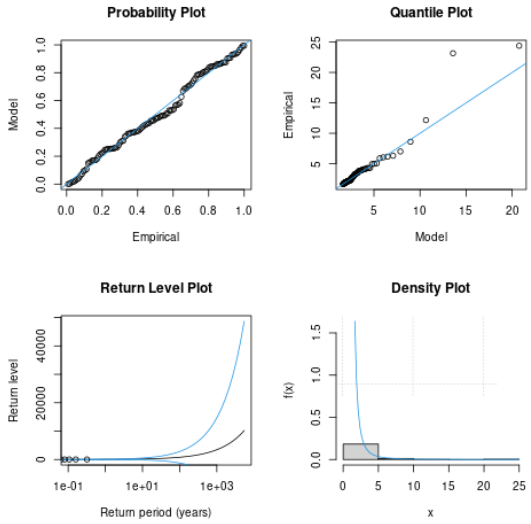
[10] Xue Liu et al. “Droughts in Wind and Solar Power: Assessing Climate Model Simulations for a Net-Zero Energy Future”. In: *Geophysical Research Letters* 51.24 (2024), e2024GL109416. DOI: https://doi.org/10.1029/2024GL109416. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL109416.

[11] Dev Millstein et al. “Wind energy variability and links to regional and synoptic scale weather”. In: *Climate Dynamics* 52 (Apr. 2019). DOI: 10.1007/s00382-018-4421-y.

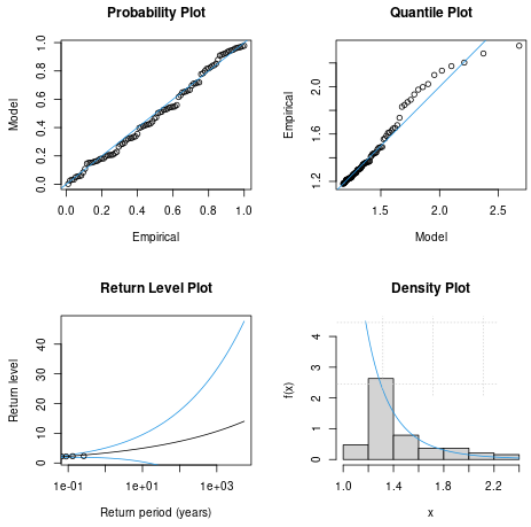
[12] Hans Van de Vyver. “Fast generation of high-dimensional spatial extremes”. In: *Weather and Climate Extremes* 46 (2024), p. 100732. ISSN: 2212-0947. DOI: https://doi.org/10.1016/j.wace.2024.100732. URL: https://www.sciencedirect.com/science/article/pii/S2212094724000938.

Appendices

A Diagnostic plots

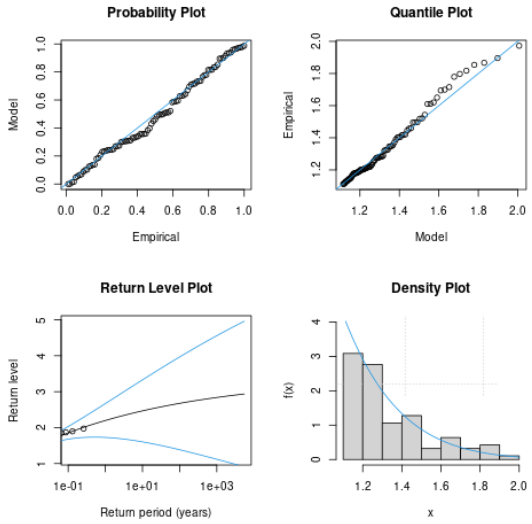


(a) $\delta = 0$

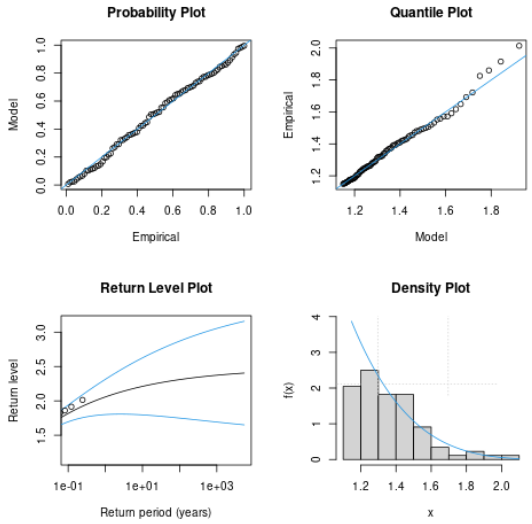


(b) $\delta = 0.25$

Figure 7: Diagnostic plots (for different combinations of the energy mix (1)) for the generalized Pareto distribution fit to the transformed declusterd Ukkel data.

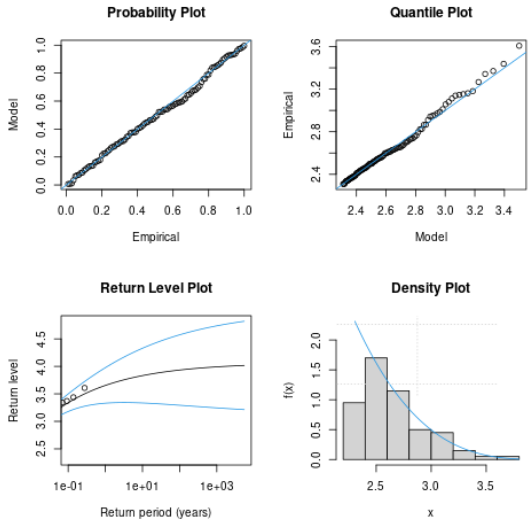


(a) $\delta = 0.5$



(b) $\delta = 0.75$

Figure 8: Diagnostic plots (for different combinations of the energy mix (1)) for the generalized Pareto distribution fit to the transformed declusterd Ukkel data.



(a) $\delta = 1$

Figure 9: Diagnostic plots (for different combinations of the energy mix (1)) for the generalized Pareto distribution fit to the transformed declusterd Ukkel data.

B Data transformation

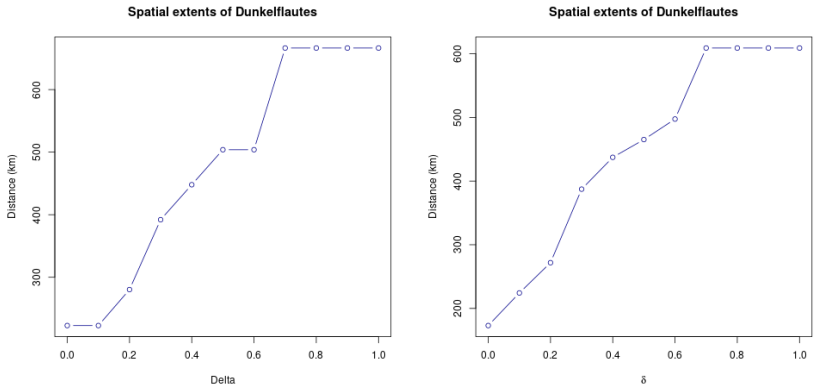
The exact transformation of the spatiotemporal data $((x_{t,1}^{orig}, \ldots, x_{t,d}^{orig}))_{t=1}^{m}$ for each location r_i is given by

$$x_{t,i} = [1 - \hat{F}_i(-x_{t,i}^{orig})]^{-1/\alpha}, \quad t = 1, \ldots, m, \tag{5}$$

where $\alpha = 2$. In order to prevent the expression on the r.h.s. of (5) to become undefined, $\hat{F}_i(x)$ is a modified empirical cumulative distribution function, given by

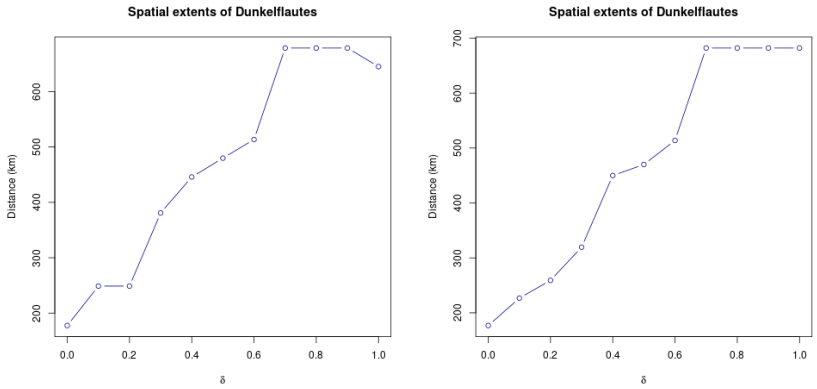
$$\hat{F}_i(x) = \frac{1}{m+1} \sum_{t=1}^{m} \mathbb{1}\{-x_{t,1}^{orig} \leq x\}.$$

C Spatial extent variation



(a) Spatial extent for 15 bins

(b) Spatial extent for 20 bins



(c) Spatial extent for 25 bins

(d) Spatial extent for 30 bins

Figure 10: Different typical spatial extents are observed, for varying number of distance bins. This is a source of uncertainty.