Nat. Hazards Earth Syst. Sci. Discuss., 3, 7457–7486, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/7457/2015/
doi:10.5194/nhessd-3-7457-2015

Natural Hazards and Earth System Sciences Discussions
Open Access

The observed clustering of damaging extra-tropical cyclones in Europe

S. Cusack

Risk Management Solutions, Peninsular House, 30 Monument Street, London, EC3R 8NB, UK

Received: 12 November 2015 – Accepted: 10 December 2015 – Published:
18 December 2015

Correspondence to: S. Cusack (stephen.cusack@rms.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The clustering of severe European windstorms on annual timescales has substantial impacts on the re/insurance industry. Management of the risk is impaired by large uncertainties in estimates of clustering from historical storm datasets typically covering the past few decades. The uncertainties are unusually large because clustering depends on the variance of storm counts.

Eight storm datasets are gathered for analysis in this study in order to reduce these uncertainties. Six of the datasets contain more than 100 years of severe storm information to reduce sampling errors, and the diversity of information sources and analysis methods between datasets sample observational errors. All storm severity measures used in this study reflect damage, to suit re/insurance applications.

It is found that the shortest storm dataset of 42 years in length provides estimates of clustering with very large sampling and observational errors. The dataset does provide some useful information: indications of stronger clustering for more severe storms, particularly for southern countries off the main storm track. However, substantially different results are produced by removal of one stormy season, 1989/1990, which illustrates the large uncertainties from a 42-year dataset. The extended storm records place 1989/1990 into a much longer historical context to produce more robust estimates of clustering. All the extended storm datasets show a greater degree of clustering with increasing storm severity and suggest clustering of severe storms is much more material than weaker storms. Further, they contain signs of stronger clustering in areas off the main storm track, and weaker clustering for smaller-sized areas, though these signals are smaller than uncertainties in actual values. Both the improvement of existing storm records and development of new historical storm datasets would help to improve management of this risk.

1 Introduction

Mailier et al. (2006) analysed the NCEP reanalysis dataset (Kalnay et al. 1996) and found clustering of winter wind storm occurrences in Europe, with some evidence that the clustering may be stronger for more severe storms. An analysis of similar data by Vitolo et al. (2009), and of other reanalysis datasets by Pinto et al. (2013), found similar results, and supplied clearer evidence of stronger clustering of the more severe storms.

The most important practical issue caused by significant clustering of severe storms is the threat to the solvency of re/insurance companies. The first step towards a more robust re/insurance industry, one which can better withstand extreme annual losses, is to measure the observed annual clustering of storms for different severities. Meteorological measures of storm severity are common in published work, such as relative vorticity at 850 hPa used by Mailier et al. (2006) and Vitolo et al. (2009), or the depth of the central pressure used by Pinto et al. (2013) and Economou et al. (2015). The damage potential of these storms is a more appropriate measure of storm severity for insurance purposes, taking into account its variability with local wind climate (Klawa and Ulbrich, 2003), and will be used throughout this study to characterise storm strength.

Karremann et al. (2014a) used severity metrics which were validated for re/insurance purposes, and measured storm severity in terms of local return levels. This use of standard insurance industry expressions of severity makes their results more relevant to end-users, but perhaps more importantly, all storm severity metrics can be easily translated to this common scale of return levels to enable inter-comparison of disparate severity measures. Return levels will be used in this study to allow inter-comparison of a wide variety of storm datasets. Karremann et al. (2014b) extend results from Germany to many other countries impacted by wind storms to provide a fuller picture of clustering as a function of local storm severity in Europe. However, the true clustering climate is obscured by large uncertainties due to sampling errors, as illustrated by the 90 % bootstrap confidence interval (CI) in Fig. 6 of Vitolo et al. (2009), based on 50 years of data. Those results imply a very wide range of true, underlying climates of storm clustering could produce the 30-year sample data of severe storms analysed by Karremann et al. (2014a, b).

The uncertainties in estimates from standard dataset samples are particularly large because clustering measures the variance of annual storm counts, rather than just the mean behaviour, leading to a greater sensitivity to errors in storm records. The large impacts of sampling and observational errors on measures of clustering significantly reduce our knowledge of clustering from standard multi-decadal storm datasets. In Sect. 2, seven extended records of historical storms are described, in addition to a more standard dataset of 42 years in length. The seven extended historical records reduce sampling errors by their increased length, and also give insight into impacts of observational errors, because the unrelated data sources and methods of making each dataset ensure independent observational errors.

Section 3 describes the method of analysing data and has two main parts: first, the measure of clustering for a group of storms is defined, and second, the method of converting the disparate measures of storm severity in the eight different datasets to a common form is described. The observed clustering of European windstorms is presented in Sect. 4, together with a discussion of estimates and errors. A summary is given in the final section.

2 Data

A total of eight storm datasets are used in this study, all of which contain the date and a measure of damage severity of each storm.

Two extended datasets of storms in the UK are studied. The first is the list of storms and their Storm Severity Index (SSI) values listed in pages 8 to 10 of Lamb and Frydendahl (1991). Their SSI measures are estimated from surface weather reports, meteorological analyses and damage information from a variety of documentary sources and reflect the damage severity of storms. The clustering analysis presented in Sect. 4 is restricted to the storms in the period from 1690 to 1989, due to incompleteness of reportage in earlier times, and to those 44 storms with SSI values of 2000 or higher. This high severity threshold ensures both a more homogeneous time-series and more confident estimates of their severity, due to the increased attention and better documentation of the most severe storms in this period.

The second UK dataset is a list of storm fatalities in the UK in the period 1835 to 1994 gathered by Risk Management Solutions (hereafter RMS). This was extracted from archives of The Times newspaper by searching its Index using the terms "storm" and "gale" (Robert Muir-Wood, personal communication, 2015). The fatalities are considered to be accurately reported throughout this period, and the dataset is considered complete since bigger national issues would reduce space or prominence attached to more minor storms events but not remove them completely. Two factors were applied to reported fatalities to homogenise this dataset: first, a population factor indexes all fatalities to 1994 national population levels, and second, night-time storm fatalities are scaled by a factor four to produce as-if daytime fatalities. Storm fatalities reflect population densities hence this index is more closely related to actual damage than wind speed intensity, and given the much more densely populated southern half of the UK, the dataset is viewed as an index of storm damage severity in the southern half of the country. Figure 1 shows a time-series of standardised storm fatalities for the full 160 year record.

Extended 105-year records of winds at five stations from the Royal Netherlands Meteorological Institute (KNMI) are used to define storminess in the Netherlands in the period from 1910 to 2014. The data and analysis are described in Cusack (2013). In brief, the winds from the five stations are merged to form an aggregate SSI value for each storm. The data are complete, and the spread of station locations geographically ensures the storm severity represents national values. The largest uncertainties arise from several significant changes in wind measurement practice in the first few decades. Intensive homogenisation methods are applied, based on station metadata made available by KNMI, complemented with statistical methods (see Supplementary Information of Cusack, 2013). The homogenisation serves to reduce but cannot completely remove observational errors, and the final time-series of storm severities will inevitably contain uncertainties. The top 30 or so storms have been compared with documentary sources such as the KNMI list available at http://projects.knmi.nl/hydra/cgi-bin/storm_list.cgi, and other independent sources based on documentary records, and corroborate the significant storms in this KNMI-derived dataset.

The public website of Deutscher Wetterdienst (DWD) provides peak gust data and associated metadata for climate stations covering the past 60 years. Seven stations with minimal changes to the wind observing system over their entire records were chosen, with locations shown in Fig. 2. SSI values for Germany were computed for individual storms over the past 60 years using the method from Cusack (2013) applied to these seven stations. While the stationary observational practices reduce uncertainties in results from inhomogeneities, the small number of selected stations covering such a large area introduces errors in estimated severity. The top storms produced by this analysis were compared to the list of DWD storms provided in Table 1 of Karremann et al. (2014a) – based on much higher station density – and there is high correlation. The larger spatial extent of more severe storms leads to this result.

Brázdil et al. (2004) describe windstorm damage in the Czech Republic from 1500 to 1999 based on research of a wide variety of documentary sources. Their detailed descriptions have been manually analysed into two storm severity classes: class 1 for local-scale damage, or large-scale weak damage, and class 2 for widespread, intense damage. Summer storms forced by convection have been removed. Figure 3 displays the number of storms per century for each severity class. Strong temporal trends can be seen in these data: there is a large increase in frequency of weaker storms in the last 200 years, and increasing occurrence of the stronger storms throughout the period. These temporal trends are most likely due to changes in amount of documentary evidence through time. Figure 3 indicates that the reduction in sampling error achieved by such a long dataset will be offset to some extent by larger uncertainties from reporting inhomogeneities. The impact of these non-stationarities will be explored in the Results section.

Stucki et al. (2014) describe a database of wind storms in Switzerland during the period 1859–2011. In brief, they use a wide variety of information, including damage information from buildings and forestry and meteorological information from anemometers and reanalyses, to identify storm events then assign one of three severity ratings to each storm, depending on the severity and spatial scale of damage in Switzerland. Summer wind storms are not infrequent in Switzerland, and all damaging wind events from May to September in the Stucki et al. (2014) database are excluded from this analysis of extra-tropical cyclone clustering. A full listing of the wind damage events in their database is given in the Supplementary Information of Stucki et al. (2014).

Emmanuel Garnier (private communication) provided a dataset of storms in France covering the period 1500–1999 based on his research of documentary archives for descriptions of wind damage. The historical storms are assigned a severity using the Beaufort scale, based on the documented damage severity and spatial extent. The present analysis will focus on the 1650 to 1999 period when documentary evidence is considered more complete and homogeneous for severe storms. Our internal validation indicated gaps in the Garnier record in the 19th century, and this has been alleviated by the inclusion of storm information from Bessemoulin (2002) to produce a more complete dataset of historical storm damage in France, though the completeness of the dataset in the late 18th and early 19th centuries is uncertain, since both sources contain little information in a period when other parts of Europe were stormy, especially in the 1815 to 1840 period. The count of storms by decade is shown in Fig. 4, split by their damage severity measured using the Beaufort Scale.

The storm footprints described in Bonazzi et al. (2012) are the most spatially comprehensive set of storms. In brief, these footprints are derived from datasets of weather station peak gusts from fifteen countries beginning in 1972. The gust datasets comprise freely available data from national meteorological services, together with some RMS purchases from private providers. Each storm footprint consists of the maximum observed gusts at each station for the entirety of the storm, which are then spatially interpolated to a more regular grid. The SSI is used to characterise the damage severity of these storms using the method described in Cusack (2013). The 135 footprints used in Bonazzi et al. (2012) are supplemented with seven storms in 2011 to 2013 (Yoda in November 2011; Friedrich, Joachim and Patrick/Dagmar in December 2011; Ulli in January 2012; Christian in October 2013 and Xaver in December 2013) using the same data and methods, to form a set spanning the 42 year period from 1972 to 2013. This set of 142 storms contains the top 20 to 25 of the strongest storms in major countries such as Germany, France, UK and Netherlands, and the top 10 to 20 storms in other countries, for the 1972-2013 period. However, due to the large footprints, all fifteen countries are affected by many more storms than these limits. More specific details on this dataset can be found in Bonazzi et al. (2012).

3 Analysis method

The strength of clustering used in most research to date adopts the metric first proposed by Mailier et al. (2006). Given a time-series of annual storm counts, X_i, where $i = 1, 2, \ldots, N$ and N is the total number of storm years, Mailier et al. (2006) defined clustering using the dispersion statistic D:

$$D = \frac{\mathrm{Var}(X)}{E(X)} - 1, \tag{1}$$

where $\mathrm{Var}(X)$ is the variance and $E(X)$ is the expected (or mean) value of observed yearly storm counts. As the variance of a Poisson process is identical to its expected value, Eq. (1) can be re-written as:

$$D = \frac{\mathrm{Var}(X) - \mathrm{Var}(\mathrm{Pois})}{\mathrm{Var}(\mathrm{Pois})}, \tag{2}$$

where Var(Pois) is the variance of a Poisson process with expected value $E(X)$. Mailier et al.'s metric of clustering is the relative excess variance of the data above a Poisson process.

Raschke (2015) analysed Eq. (1) and found D is proportional to the total rate of storms in the set being analyzed. Therefore, the metric reflects both the strength of clustering and the size of the storm group studied. Raschke proposes a new metric of clustering called “Beta” to distinguish clustering strength from size of storm group being studied. We re-name Raschke’s metric as the clustering coefficient (CC) to reduce confusion with other uses of “Beta” in research. Raschke’s metric simplifies to the dispersion statistic given by Eq. (1) normalized by the mean value of observed yearly storm counts (the rate):

$$\mathrm{CC} = \frac{D}{E(X)}. \tag{3}$$

Since the variation of clustering with storm strength will be explored, and more severe storms are rarer, Eq. (3) will be used for all results in Sect. 4 to ensure no artefact of dependence on storm numbers.

For each dataset, all storms matching or exceeding a specified damage threshold in storm years defined from July to following June were identified, then Eq. (3) was used to determine the strength of clustering. Various damage thresholds are used in each dataset to reveal the variation of clustering strength with storm severity. These severity thresholds are expressed as return levels, following Karremann et al. (2014a, b), and we refer to them as return periods (RP). In brief, the RP is defined to be the inverse of the annual frequency of storms greater than or equal to the particular threshold severity. For example, if a group of storms contain an average annual rate of 0.5 storms per year matching or exceeding the threshold, then the storm severity is defined to be RP = 2 years. This representation unifies dissimilar measures of severity (e.g. SSI, damage classes in Switzerland, UK storm fatalities) to enable their inter-comparison.

4 Results and discussion

Figure 5 displays the variation of clustering with storm severity based on the Europe-wide footprints derived from the dataset of station peak gusts covering 42 storm seasons, from 1972/1973 to 2013/2014. Results in Fig. 5 suggest greater clustering for more severe storms, consistent with published work (e.g., Mailier et al., 2006; Vitolo et al., 2009; Pinto et al., 2013).

The dashed lines in Fig. 5 represent the 95th confidence interval based on sampling error estimated using 50 000 sets of data, each of which are 42-year samples drawn from a random Negative Binomial process with parameters fitted to each data point in Fig. 5. The dotted lines in Fig. 5 represent an approximate measure of the uncertainty from observational error. Its method of approximation is illustrated by example. The value at RP3 is usually based on the dates of the top 14 storms. To explore uncertainty in estimates of storm severity, the top half of the subset, 7 storms, are assumed to be fixed, then the next 14 storms are identified, and 7 storms are selected randomly from this second tier of 14 storms, which are then added to the top 7 to form a different sample of the "top" 14 storms. The clustering is measured for each of 1000 random samples, and the empirical 0.025 and 0.975 quantiles are displayed in Fig. 5. The assumptions that the stronger 50 % of the storm subset are known, and the weaker 50 % of the storm subset are uncertain, is intended to serve as a guide to the impact of observational errors. Figure 5 indicates large uncertainty in estimated CC values from both sampling and observational errors. Combining these two sources of uncertainty leads to the conclusion that the amount of clustering at any specific severity threshold would not be distinguished from a Poisson process (CC = 0) at the 5 % level.

This assessment of uncertainties at individual points is distinct from the broader question of whether the collection of all data points in Fig. 5 is clustered. This is assessed as follows: a set of storms equal to the largest rate (2.0 in Fig. 5, or RP = 0.5) is created, with randomly assigned storm strengths, then a time-series of occurrence following a random Poisson process is generated; the clustering coefficient is computed for each severity threshold, depending on the earlier designated severity assignments; this is repeated with 50 000 random sets of data, to form 50 000 Poisson samples of CC vs RP. The empirical probability that the CC of the observed storms is greater than the Poisson sample is recorded at each RP, and the probabilities at each RP are multiplied together to form a score corresponding to the likelihood that the observed CC values are above that of a Poisson process. The likelihood score is computed for each of the 50 000 Poisson samples, and it is found that the observations exceed 99.6 % of all Poisson samples. This leads to the conclusion that European storms with severity between RPs of 0.5 and 3 years are significantly different from a Poisson process at the 1 % level.

The question of whether there is an increase in the clustering for more severe storms is now addressed by analysing the same data used in Fig. 5. The method is as follows: compute the best linear fit between observed CC and severity expressed as the logarithm of RP; fit Negative Binomial model parameters to observed time-series at RP = 0.5 threshold; generate a random Negative Binomial sample and assign storm strength ranks randomly to it, then form subsets for each RP severity threshold (this is essentially the same method as above, except for a Negative Binomial rather than a Poisson); compute CC vs RP for this random sample, then find the best fitting gradient of CC vs log(RP); finally, repeat this 50 000 times to obtain a set of 50 000 gradients. It was found that the gradient of CC versus severity in the observed storm set was more positive than 98.9 % of all randomly generated samples. This leads to the conclusion that greater clustering with stronger storms at the Europe-scale is much more likely than not, though the fact that more than 1 % of samples with randomly assigned severity relationships have a more positive gradient indicates some uncertainty in this finding.

The clustering behaviour at national scales is now investigated using the RMS 42-year storm footprint dataset. Figure 6a displays CC versus RP curves for countries in the northern part of the European windstorm affected area, while Fig. 6b shows curves for more southern countries. The large uncertainties depicted in Fig. 5 apply to national scales too (not shown). Thus the differences between northern countries in Fig. 6a lie well within the limits of error, and similarly for southern countries in Fig. 6b. However, comparison of Fig. 6a and b reveals a signal of stronger clustering for more severe storms in the southern part of the domain. The main driver of this north-south difference is the exceptional nature of the storminess in January to March 1990 in the southern countries. Figure 7 contains CC versus RP curves for southern countries when the 1989/1990 storm season is removed, and it can be seen how clustering strengths at RPs of 1 to 3 years are now much more similar between northern (Fig. 6a) and southern (Fig. 7) parts of the domain. This exemplifies the large sampling errors shown in Fig. 5: if this season had not occurred, the clustering strengths in more southern countries would be very different (Fig. 7 versus 6b). The conclusion is that sampling errors are of major importance when studying the past few decades. Longer records would help to reduce such large sampling errors and place 1989/1990 into a fuller historical context. This is the motivation for analysing longer historical datasets.

Figure 8 contains results from an analysis of the longer storm datasets in the UK and Netherlands. Figure 6a indicates low values of CC in the UK at all RPs, and a test of the hypothesis that the group of all data points are significantly different from a sample of Poisson data is rejected at the 0.1 significance level, in common with most northern countries. The results from extended UK storm datasets in Fig. 8a show CC values of about 1.0 for storms with severities exceeding RPs of 5 years. The extended lengths of the two UK datasets and their independent methods of gathering and assessing storm severities point to significantly smaller uncertainties than those shown in Fig. 5, raising confidence that more severe storms in UK are clustered. The analysis of the 105-year storm dataset in the Netherlands (Fig. 8b) is consistent with the low levels of clustering found in the 42-year RMS dataset. The raised clustering value at the RP of 6 years in the longer Dutch storm dataset is very uncertain due to limited sample sizes. However, similar behaviour from two longer and entirely independent datasets in the neighbouring UK supports the raised clustering of storms above RP6 severity in the extended Dutch storms dataset.

Figure 9 contains the clustering strengths found in four extended datasets in the southern part of the study area. Results in Fig. 9a indicate lower levels of clustering in the 59-year DWD storm dataset compared to the 42-year RMS datasets. The DWD clustering is more similar to the 42-year RMS dataset with 1989/1990 removed. This may be due to greater weighting of far northern Germany in the DWD dataset (3 of the 7 stations), since the 1989/1990 season was less extreme in this area, relative to local storm climate. The dotted lines in Fig. 9a represent CC versus RP when one station is removed from the DWD dataset, and demonstrate how DWD clustering is not especially sensitive to any single weather station.

The results of analysing the 350-year Garnier-Bessemoulin dataset are shown alongside the 42-year dataset in Fig. 9b. The very long storm dataset contains clear signs of clustering of the most severe storms in France. The independence of the information sources, and the increased length of the Garnier-Bessemoulin dataset, raises confidence in the finding of stronger clustering of more severe storms in the RMS dataset.

Figure 9c shows the results from analysing the very long storm dataset in Czech Republic from Brázdil et al. (2004), alongside those from the 42-year RMS wind datasets. The results from the shorter dataset showed great sensitivity to the inclusion of the 1989/1990 storm season and an independent, longer dataset is very useful to help place 1989/1990 in historical context. However, the reporting inhomogeneities in this long dataset (Sect. 2) are a source of significant uncertainty in results. Table 1 shows the clustering coefficient for class 1 storms for a range of different time periods in the Brázdil dataset, and Table 2 shows results for class 2 storms. CC varies substantially according to the time period studied, though a clear signal emerges of lower values at RP threshold of around 1 year, and significantly stronger clustering of more severe storms (RP threshold of around 10 years). Using the information in Fig. 3, the 1800 to 1999 period is chosen to represent clustering of class 1 storms and stronger, whereas 1700 to 1999 is chose to represent class 2 storms, in Fig. 9c. The main finding from this much longer dataset is weaker clustering around RP1 thresholds and notably stronger clustering of more severe storms. Further investigation of the 42-year dataset at shorter RP thresholds reveals a 6 year period of elevated gust readings from about 1989 to 1995 suggesting inhomogeneous observation practices. This adds to the acute sensitivity of CC to the inclusion of the 1989/1990 season in the shorter dataset, as shown in Fig. 9c. The existence of significant observational errors in the records of more recent storms illustrates the benefits of analysing multiple, independent storm datasets.

Figure 9d contains the results from an analysis of Swiss storms. The extended storm damage dataset from Stucki et al. (2014) indicates weak clustering at shorter RPs, and slightly larger values at longer RPs, which supports the findings from the 42-year dataset. CC values are lower than in nearby France, Germany and Czech Republic around RP1 to 3 thresholds. The most unique feature of Switzerland relative to these nearby countries is its much smaller spatial extent. This suggests a dependence of local CC values on size of area studied, which is consistent with the lower dispersion values for narrower latitudinal barriers reported in Vitolo et al. (2009).

Results from all extended storm datasets are presented in Fig. 10. The results contain two main features. First, there is generally stronger clustering in southern countries: at shorter RPs, the Netherlands CC values are generally below those of Germany, Czech Republic and Switzerland, while the UK values at longer RPs are generally lower than in France and the Czech Republic. This geographical variation is consistent with that found by comparing Fig. 6a and b, however, the signal is smaller in longer datasets. Given the varied nature and independence of these datasets, and their much longer records of storm history, they add weight to the argument that countries further from the main storm track in Europe experience stronger clustering of storms, though the difference is small relative to uncertainties in knowledge of clustering in any country. The second notable aspect of results in Fig. 10 concerns the earlier finding of a strong sensitivity of CC values in more southern countries to inclusion of the 1989/1990 season (Figs. 6b and 7). The CC values around RP1 to 3-year thresholds from the extended datasets are lower than those in Fig. 6b (with 1989/1990) and closer to those in Fig. 7 (without 1989/1990). This is a practical illustration of large sampling errors in shorter

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀| |▶
◀ ▶
Back Close
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

datasets: too much weight is placed on the big cluster in 1989/1990 inflating CC values, and longer-term records place the 1989/1990 storm cluster in fuller historical context.

5 Summary

The clustering of extra-tropical cyclones in Europe has been investigated from the perspective of the re/insurance sector since they suffer the most material impacts from this phenomenon. Specifically, storm severities are measured in terms of damage potential, storms are gathered into groups according to exceedance of damage severity thresholds expressed as return periods (RP), and clustering on annual timescales is studied.

Perhaps the most notable characteristic of clustering is the unusually large uncertainties of estimates based on typical storm dataset lengths of a few decades, due to its dependence on storm count variance. This was found in previous research and has been explored in more detail in this study. Both the sampling and observational errors are large and dominate estimates of clustering for any single group of storms.

Eight different storm datasets were gathered to reduce these large uncertainties in estimates. The mix of different information sources and storm severity measures give a guide to observational errors, and six of the datasets were more than 100 years in length and serve to reduce sampling errors. Quality control was applied to each dataset: the biggest issue with such long datasets is temporal inhomogeneity and the period of analysis was shortened for some datasets to improve the quality of the analysis. Finally, the inter-comparison of data with different units of storm severity (e.g. SSI, damage severity classes, fatalities) was made possible by expressing each dataset's storm severities in units of local RP.

The evidence from all datasets strongly suggests that clustering increases with storm severity, from RPs of 0.5 years out to 10 or more years. The 42-year RMS storm database shows a distinction between northern areas with weaker clustering to southern parts with stronger clustering of severe storms, however, the removal of one very stormy season (1989/1990) eliminates differences between the two regions. This epitomises the large sampling errors of clustering estimates based on a few decades of data. The longer datasets contain signs of stronger clustering in countries off the main storm track, with notable years in history of multiple severe storms. Conversely, countries closer to the storm track show little signs of clustering of storms at RPs around one year, though three longer datasets in the UK and Netherlands indicate some clustering of more severe storms. While the consistency of the signal across multiple datasets increases the likelihood of this being a real feature, the difference between regions is small compared to uncertainties in clustering strength in any particular country. Finally, the comparison of clustering in Switzerland with larger neighbours indicates weaker clustering with smaller spatial scales of analysis, which is consistent with earlier published findings.

While the multiple datasets used in this study reduce uncertainties in estimates of the clustering of severe windstorms, there is plenty of room for further reductions. Europe is relatively rich in historical documentation and expanded research into these archives would help to more accurately define the risk from windstorm clustering.

Acknowledgements. The author would like to thank Prof. Brázdil, Peter Stucki and all their collaborators for making their storm data publicly available, and more generally to KNMI and DWD as well. This study benefitted from discussions with Robert Muir-Wood, Steve Jewson and Christos Mitas.

References

Bessemoulin, P.: Les tempêtes en France, Annales des Mines, Réalités Industrielles, 9–14, 2002.

Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S.: The spatial structure of European wind storms as characterized by bivariate extreme-value Copulas, Nat. Hazards Earth Syst. Sci., 12, 1769–1782, doi:10.5194/nhess-12-1769-2012, 2012.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀▌ | ▶▌
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Brázdil, R., Dobrovolny, P., Stekl, J., Kotyza, O., Valasek, H., Jaroslav, J., History of weather and climate in the Czech lands VI: Strong winds, Masaryk University, Brno, Czech Republic, 377 pp,. 2004.

Cusack, S.: A 101 year record of windstorms in the Netherlands, Climatic Change, 116, 693–704, 2013.

Economou, T., Stephenson, D. B., Pinto, J. G., Shaffrey, L. C., and Zappa, G.: Serial clustering of extratropical cyclones in a multi-model ensemble of historical and future simulations, Q. J. Roy. Meteorol. Soc., 141, 3076–3087, doi:10.1002/qj.2591, 2015.

Kalnay, E., Kanamitsu, M., Kirtler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha ,S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, 1996.

Karremann, M. K., Pinto, J. G., von Bomhard, P. J., and Klawa, M.: On the clustering of winter storm loss events over Germany, Nat. Hazards Earth Syst. Sci., 14, 2041-2052, doi:10.5194/nhess-14-2041-2014, 2014a.

Karremann, M. K., Pinto, J. G., Reyers, M., and Klawa, M.: Return periods of losses associated with European windstorm series in a changing climate, Environ. Res. Lett., 9, 124016, doi:10.1088/1748-9326/9/12/124016124016, 2014b.

Klawa, M. and Ulbrich, U.: A model for the estimation of storm losses and the identification of severe winter storms in Germany, Nat. Hazards Earth Syst. Sci., 3, 725–732, doi:10.5194/nhess-3-725-2003, 2003.

Lamb, H., and Frydendahl, K., Historic storms of the North Sea, British Isles, and Northwest Europe, Cambridge University Press, Cambridge, 204 pp., 1991.

Mailier, P. J., Stephenson, D. B., Ferro, C. A. T., and Hodges, K. I.: Serial clustering of extratropical cyclones, Mon. Weather Rev., 134, 2224–2240, 2006.

Pinto, J. G., Bellenbaum, N., Karremann, M. K., and Della-Marta, P. M., 2013, Serial clustering of extratropical cyclones over the North Atlantic and Europe under recent and future climate conditions, J. Geophys. Res.-Atmos., 118, 12476–12485, 2013.

Raschke, M.: Statistical detection and modeling of the over-dispersion of winter storm occurrence, Nat. Hazards Earth Syst. Sci., 15, 1757–1761, doi:10.5194/nhess-15-1757-2015, 2015.

Stucki, P., Brönnimann, S., Martius, O., Welker, C., Imhof, M., von Wattenwyl, N., and Philipp, N.: A catalog of high-impact windstorms in Switzerland since 1859, Nat. Hazards Earth Syst. Sci., 14, 2867–2882, doi:10.5194/nhess-14-2867-2014, 2014.

Vitolo, R., D. B. Stephenson, D. B., Cook, I. M., and Mitchell-Wallace, K.: Serial clustering of intense European storms, Meteorol. Z., 18, 411–424, 2009.

Table 1. The CC for class 1 storms in the Brazdil dataset, for various time periods.

	1500–1599	1600–1699	1700–1799	1800–1899	1900–1999	1800–1999	1500–1999
RP (years)	1.80	2.38	1.54	0.90	0.58	0.71	1.12
CC	1.19	0.27	−0.32	0.52	0.11	0.28	0.59

Table 2. As Table 1, for class 2 storms.

	1700–1849	1850–1999	1700–1999	1500–1999
RP (years)	7.89	10.00	8.82	11.09
CC	4.07	1.74	3.19	2.97

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◂◂
▸▸
◂
▸
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

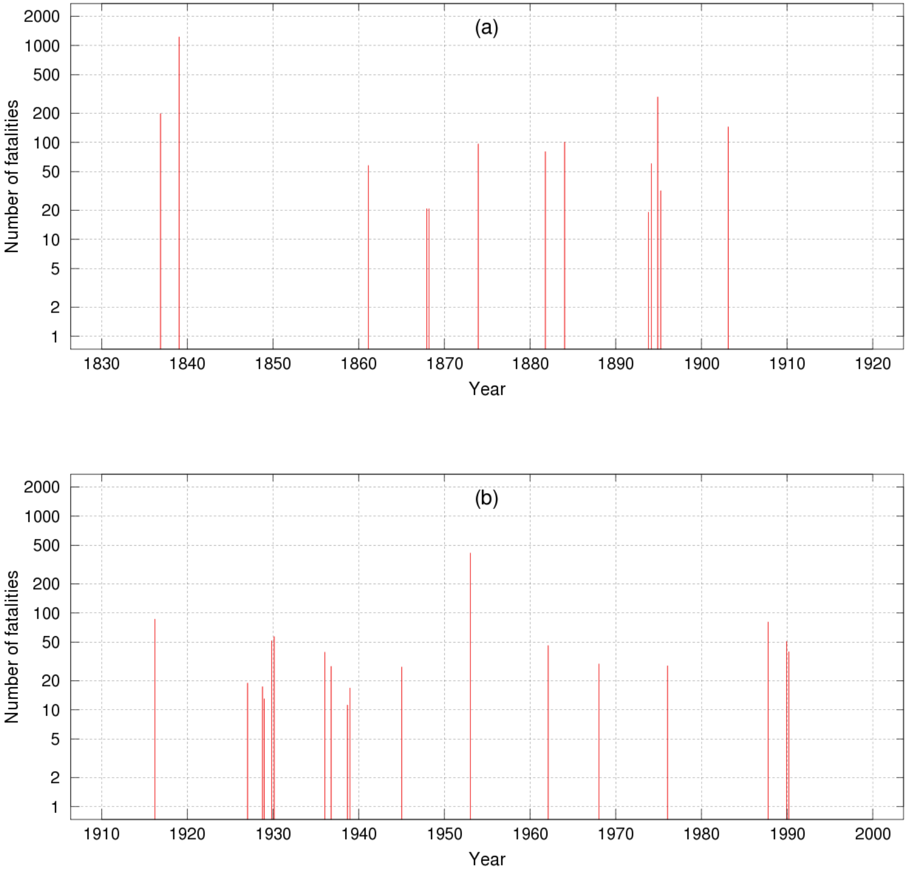

Figure 1. Time-series of storm fatalities in the UK from the RMS dataset. All data are adjusted as-if storms occurred during daytime, and trended to 1994 population levels.

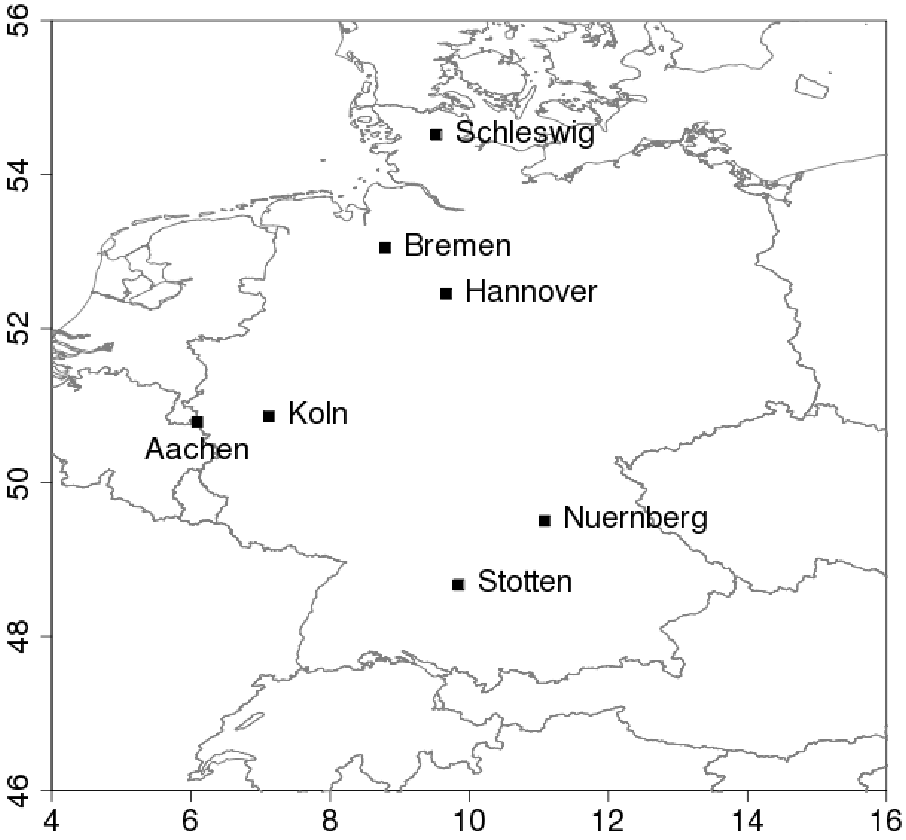

Figure 2. The location of the seven DWD weather stations in Germany.

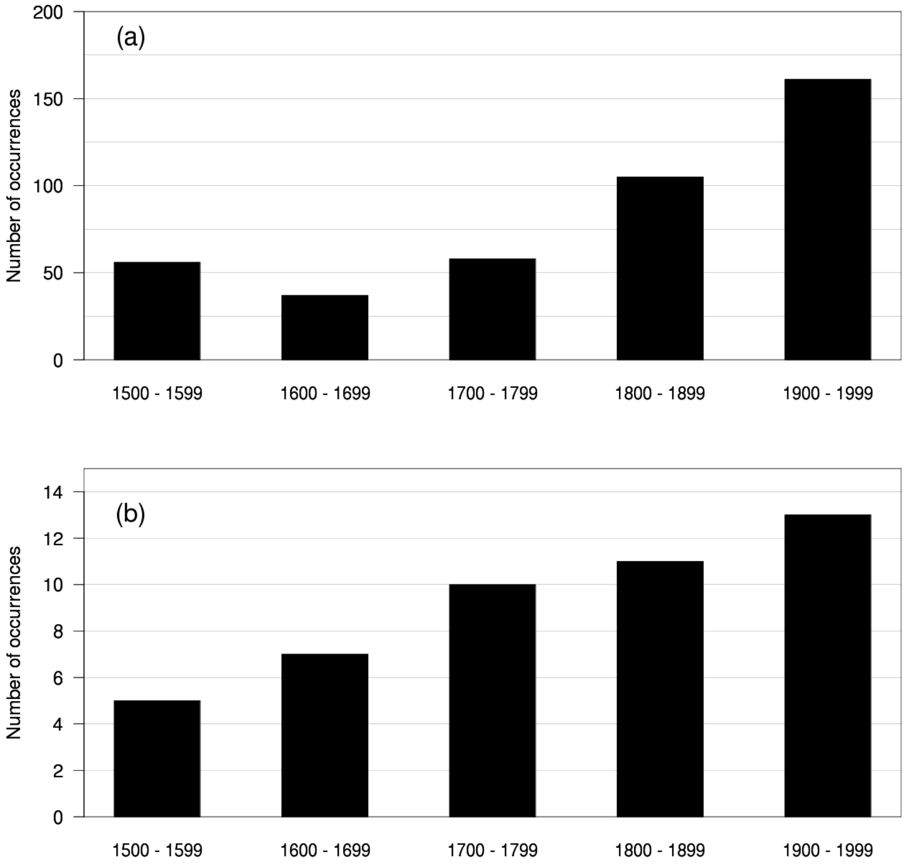

Figure 3. Histogram of storm occurrences per century in Czech Republic for **(a)** weaker class 1, and **(b)** stronger class 2 storms.

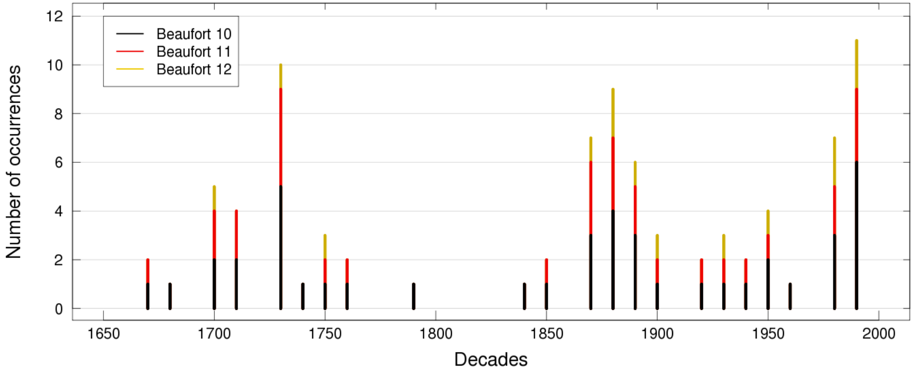

Figure 4. Count of storm occurrences per decade in France from the Garnier-Bessemoulin dataset, split into three damage severity categories.

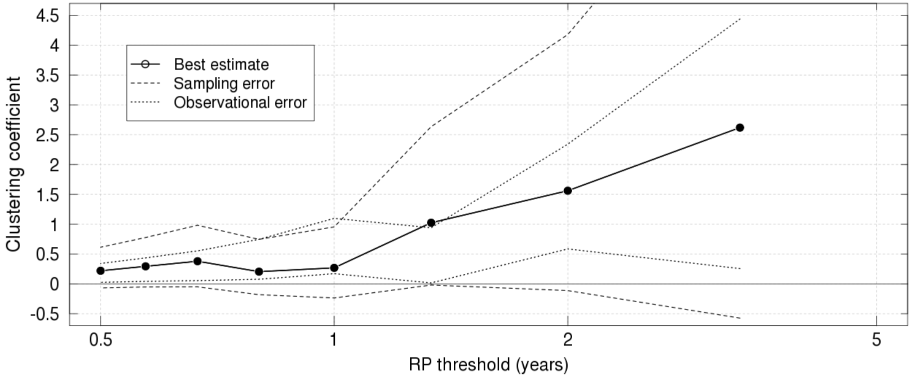

Figure 5. Clustering as a function of the storm severity groupings for historical storms in the period 1972/1973 to 2013/2014. The dashed lines show the 95th confidence interval based on sampling error, and the dotted lines represent the 95th confidence interval of observational errors.

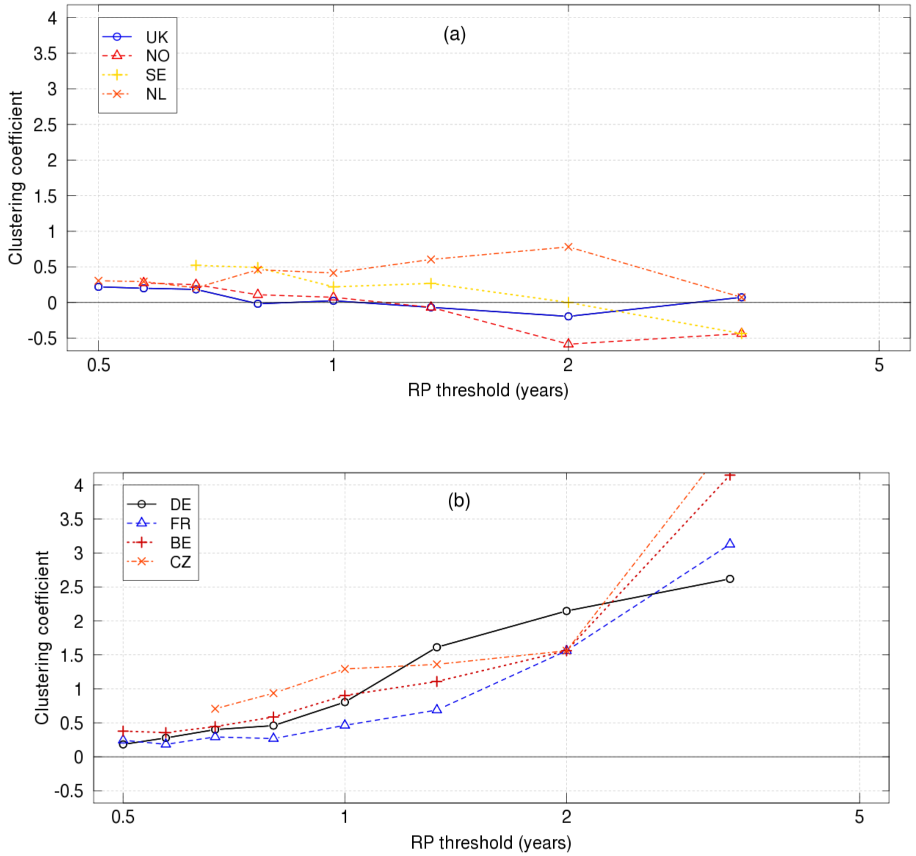

Figure 6. As Fig. 5, for various countries in **(a)** northern part and **(b)** southern part of the study area.

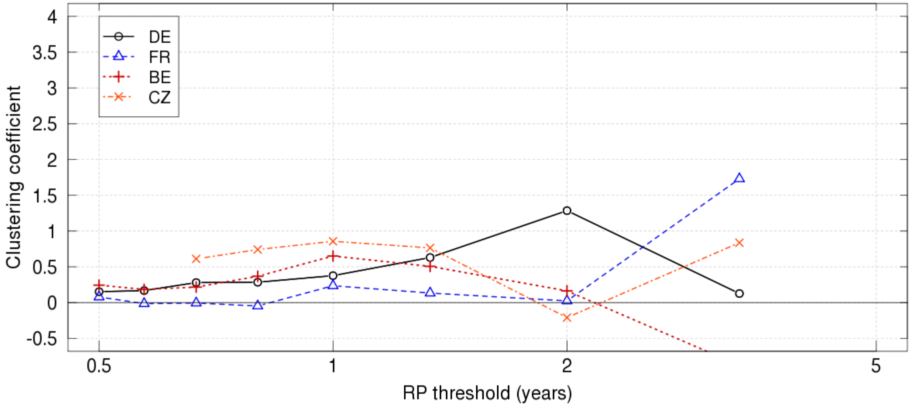

Figure 7. As Fig. 6b, with 1989/1990 storm season excluded.

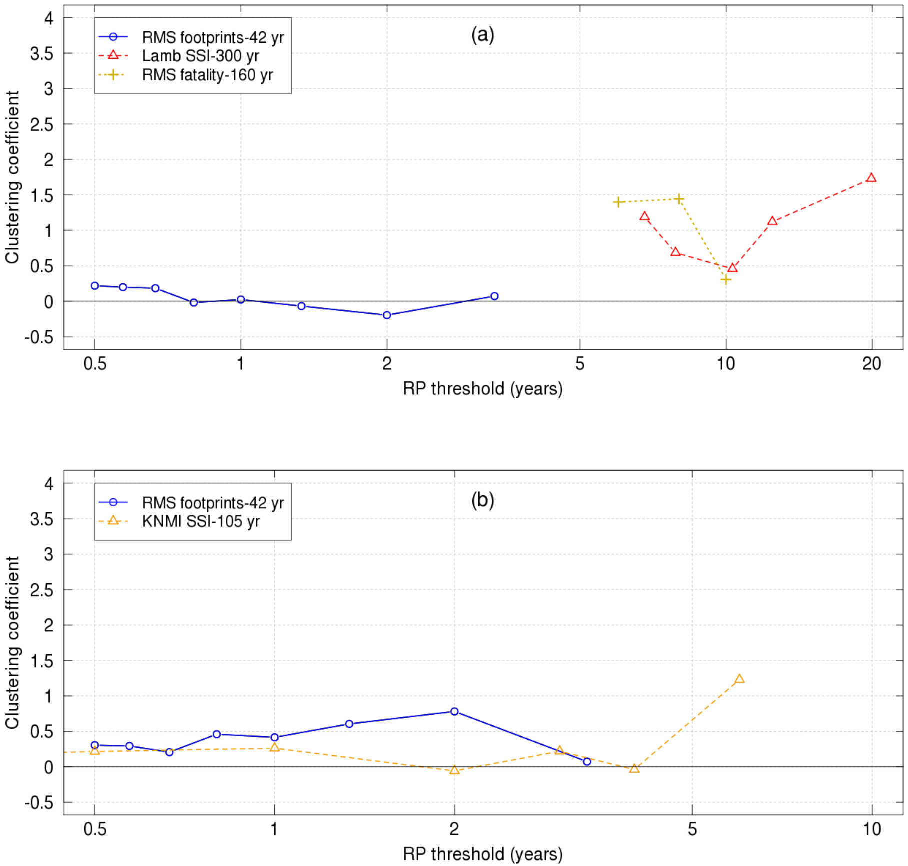

Figure 8. Clustering as a function of the storm severity groupings for **(a)** UK and **(b)** Netherlands.

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◂◂
▸▸
◂
▸
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

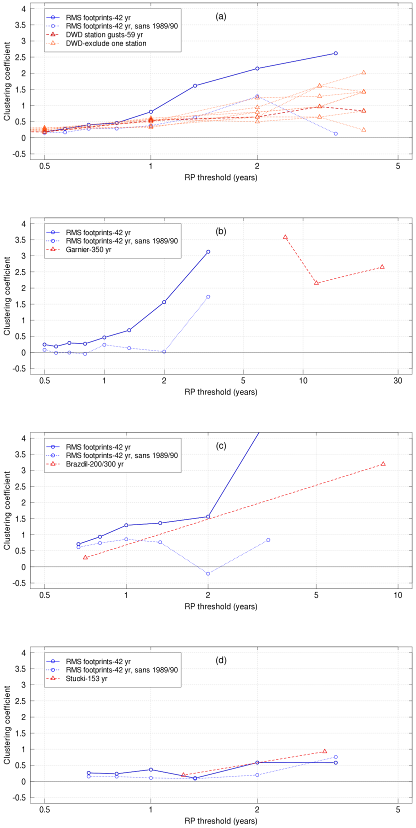

Figure 9. Clustering as a function of storm severity for **(a)** Germany, **(b)** France, **(c)** Czech Republic and **(d)** Switzerland.

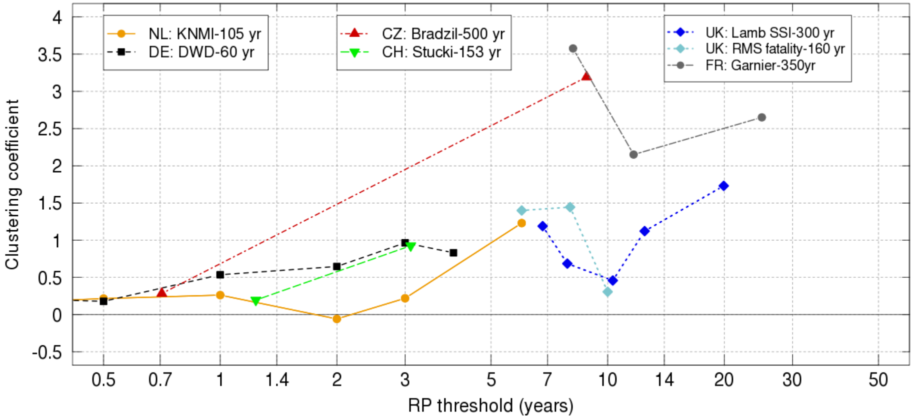

Figure 10. Clustering versus storm severity from extended historical storm datasets.